

Hyperspectral Unmixing via Turbo Bilinear Approximate Message Passing

Jeremy Vila, *Student Member, IEEE*, Philip Schniter, *Fellow, IEEE* and Joseph Meola, *Member, IEEE*

Abstract—The goal of hyperspectral unmixing is to decompose an electromagnetic spectral dataset measured over M spectral bands and T pixels into N constituent material spectra (or “endmembers”) with corresponding spatial abundances. In this paper, we propose a novel approach to hyperspectral unmixing based on loopy belief propagation (BP) that enables the exploitation of spectral coherence in the endmembers and spatial coherence in the abundances. In particular, we partition the factor graph into spectral coherence, spatial coherence, and bilinear subgraphs, and pass messages between them using a “turbo” approach. To perform message passing within the bilinear subgraph, we employ the bilinear generalized approximate message passing algorithm (BiG-AMP), a recently proposed belief-propagation-based approach to matrix factorization. Furthermore, we propose an expectation-maximization (EM) strategy to tune the prior parameters and a model-order selection strategy to select the number of materials N . Numerical experiments conducted with both synthetic and real-world data show favorable unmixing performance relative to existing methods.

Index Terms—hyperspectral imaging, approximate message passing, belief propagation, expectation-maximization algorithms

I. INTRODUCTION

In *hyperspectral unmixing* (HU), the objective is to jointly estimate the spectral signatures and per-pixel abundances of the N materials present in a scene, given measurements across M spectral bands at each of $T = T_1 \times T_2$ pixels. Often, a *linear mixing model* [2], [3] is assumed, in which case the measurements $\mathbf{Y} \in \mathbb{R}^{M \times T}$ are modeled as

$$\mathbf{Y} = \mathbf{S}\mathbf{A} + \mathbf{W}, \quad (1)$$

where the n th column of $\mathbf{S} \in \mathbb{R}_+^{M \times N}$ represents the spectrum (or “endmember”) of the n th material, the n th row of $\mathbf{A} \in \mathbb{R}_+^{N \times T}$ represents the spatial abundance of the n th material, and \mathbf{W} represents noise. Both \mathbf{S} and \mathbf{A} must contain only non-negative (NN) elements, and each column of \mathbf{A} must obey the simplex constraint (i.e., NN and sum-to-one). Recently, nonlinear mixing models have also been considered (e.g., [4], [5]), although such models lie outside of the scope of this paper.

This work has been supported in part by NSF grants IIP-0968910, CCF-1018368, and CCF-1218754, and by an allocation of computing time from the Ohio Supercomputer Center. Portions of this work were presented at the 2013 SPIE Defense, Security, and Sensing symposium [1].

J. Vila and P. Schniter are with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (vila.2@osu.edu; schniter@ece.osu.edu; phone 614.247.6488; fax 614.292.7596).

J. Meola is with the Multispectral Sensing Division, Air Force Research Laboratory, Wright-Patterson AFB, OH 45433 USA (joseph.meola@wpafb.af.mil).

Traditionally, hyperspectral unmixing is a two-step procedure, consisting of *endmember extraction* (EE) to recover the endmembers followed by *inversion* to recover the abundances. Many EE algorithms leverage the “*pure pixel*” assumption: for each material, there exists at least one observed pixel containing only that material (i.e., all columns of the $N \times N$ identity matrix can be found among the columns of \mathbf{A}). Well-known examples of pure-pixel-based EE algorithms include N-FINDR [6] and VCA [7]. The existence of pure pixels in HU is equivalent to “*separability*” in the problem of non-negative matrix factorization (NMF), where the goal is to find $\mathbf{S} \in \mathbb{R}_+^{M \times N}$ and $\mathbf{A} \in \mathbb{R}_+^{N \times T}$ matching a given $\mathbf{Z} = \mathbf{S}\mathbf{A}$. There, separability has been shown to be sufficient for the existence of unique factorizations [8] and polynomial-time solvers [9], with a recent example being the FSNMF algorithm from [10]. In HU, however, the limited spatial-resolution of hyperspectral cameras implies that the pure-pixel assumption does not always hold in practice. With “mixed pixel” scenarios in mind, algorithms such as Minimum Volume Simplex Analysis (MVSA) [11] and Minimum Volume Enclosing Simplex (MVES) [12] attempt to find the minimum-volume simplex that contains the data \mathbf{Y} .

In the inversion step, the extracted endmembers in $\hat{\mathbf{S}}$ are used to recover the simplex-constrained abundances in \mathbf{A} . Often this is done by solving [13], [14]

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A} \geq 0} \|\mathbf{Y} - \hat{\mathbf{S}}\mathbf{A}\|_F^2 \text{ s.t. } \mathbf{1}_N^\top \mathbf{A} = \mathbf{1}_T^\top, \quad (2)$$

where $\mathbf{1}_N$ denotes the $N \times 1$ vector of ones, which is usually referred to as *fully constrained least squares* (FCLS).

Real-world hyperspectral datasets can contain significant structure beyond non-negativity on s_{mn} and simplex constraints on $\{a_{nt}\}_{n=1}^N$. For example, the abundances $\{a_{nt}\}_{n=1}^N$ will be *sparse* if most pixels contain significant contributions from only a small subset of the N materials. Also, the abundances $\{a_{nt}\}_{t=1}^T$ will be *spatially coherent* if the presence of a material in a given pixel makes it more likely for that same material to exist in neighboring pixels. Likewise, the endmembers $\{s_{mn}\}_{m=1}^M$ will be *spectrally coherent* if the radiance values are correlated across frequency.

Various unmixing algorithms have been proposed to leverage these additional structures. For example, given an endmember estimate $\hat{\mathbf{S}}$, the SUnSAL algorithm [15] estimates sparse abundances \mathbf{A} using ℓ_1 -regularized least-squares (LS), and the SUnSAL-TV algorithm [16] adds total-variation (TV) regularization [17] to also penalize changes in abundance across neighboring pixels (i.e., to exploit spatial coherence). SUnSAL and SUnSAL-TV can be categorized as unmixing

algorithms, rather than inversion algorithms, since their ℓ_1 -regularization supports the use of large (i.e., $N > M$) and scene-independent endmember libraries for $\hat{\mathbf{S}}$. However, there are limitations on the size of the library $\hat{\mathbf{S}}$, and it can be difficult to determine suitable choices for the ℓ_1 and TV regularization weights.

Traditional NMF techniques have also been enhanced to account for spectral and spatial coherence. For instance, the $\ell_{1/2}$ NMF ($L_{1/2}$ NMF) [18] algorithm promotes sparse abundances by adding $\ell_{1/2}$ regularization to the traditional NMF formulation. The algorithms in [19]–[21] then expand on this idea by adding additional regularizations to promote coherent abundances. For example, the Substance Dependence constrained NMF (SDSNMF) [19], which was shown in [19] to perform the best out of [19]–[21], employs a sparse pixel-by-pixel weighting matrix that accounts for similarities in the abundances across in the scene. Additional coherence-promoting NMF techniques include a method based on hierarchical rank-2 decompositions [22], a method that promotes both abundance separability and coherence [23], and a piecewise spectral/spatial smoothness constrained method [24].

Bayesian approaches to hyperspectral unmixing have also been proposed. For example, the Bayesian Linear Unmixing (BLU) algorithm [25] employs priors that enforce NN constraints on the endmembers and simplex constraints on the per-pixel abundances, and returns either (approximately) minimum mean-square error (MMSE) or maximum a posteriori (MAP) estimates using Gibbs sampling. The Spatially Constrained Unmixing (SCU) [26] algorithm, an extension of BLU, furthermore exploits spatial coherence using a hierarchical Dirichlet-process prior. Both BLU and SCU have been shown to outperform N-FINDR and VCA-plus-FCLS under certain conditions [26], but at the cost of several orders-of-magnitude increase in runtime.

In this paper, we propose a novel empirical-Bayesian approach to HU that is based on loopy belief propagation (LBP) [27]. Our approach, referred to as HU turbo-AMP (HUT-AMP), simplifies the intractable task of LBP on the entire factor graph (see Fig. 1) by partitioning it into three subgraphs: one that models spectral coherence (using N Gauss-Markov chains), one that models spatial coherence (using N binary Markov Random Fields (MRFs)), and one that models the NN bilinear structure of (1). While the first two subgraphs yield inference problems that are handled efficiently by standard methods [28], [29], the third does not. Thus, to perform efficient inference on the latter subgraph, we apply the recently proposed Bilinear Generalized Approximate Message Passing (BiG-AMP) algorithm [30]. BiG-AMP can be interpreted as an extension of approximate message passing (AMP) techniques [31]–[33], originally proposed for the linear observation models that arise in compressive sensing, to bilinear models like (1). To merge BiG-AMP-based inference with Markov-chain and MRF-based inference, we leverage the “turbo AMP” approach first proposed in [34] and subsequently applied to joint channel-estimation and decoding [35], [36], compressive image retrieval [37], [38], and compressive video retrieval [39], all with state-of-the-art results. In formulating our statistical model, we treat the parameters of the prior distributions as

deterministic unknowns and estimate them from the data using the expectation-maximization (EM) algorithm, building on the NN sparse reconstruction work in [40]. As such, our approach can be classified as *empirical Bayesian* [41]. Lastly, when the number of materials N is unknown, we show how it can be accurately estimated using a classical model-order selection (MOS) strategy [42]. The resulting algorithm has the following desirable features: 1) it requires no tuning parameters, 2) it exploits both spectral and spatial coherence, and 3) it uses a computationally efficient inference procedure.

We evaluate the performance of our proposed technique, in comparison to several recently proposed methods, through a detailed numerical study that includes both synthetic and real-world datasets. The results, presented in Sec. IV, suggest that HUT-AMP yields an excellent combination of unmixing performance and computational complexity.

Regarding novel contributions to HU models, we believe that our work (first presented in [1]) is the first to use either of the following: i) Gauss-Markov chains to model spectral coherence in endmembers, ii) Bernoulli truncated-Gaussian mixtures to model abundance amplitudes. As for novel contributions to inference methodology, we believe that our work is the first to combine *any* of the following methods, and in fact we combine all four of them: i) compressed sensing with non-negative Bernoulli-Gaussian-mixture priors whose parameters are learned via EM [40], ii) turbo compressed sensing that combines AMP with Markov-chain inference and learns the parameters via EM [39], iii) turbo compressed sensing that combines AMP with MRF inference and learns the parameters via EM [38], iv) bilinear AMP [30].

Notation: For matrices, we use boldface capital letters like \mathbf{A} , and we use \mathbf{A}^\top , $\text{tr}(\mathbf{A})$, and $\|\mathbf{A}\|_F$ to denote the transpose, trace, and Frobenius norm, respectively. For vectors, we use boldface small letters like \mathbf{x} , and we use $\|\mathbf{x}\|_p = (\sum_n |x_n|^p)^{1/p}$ to denote the ℓ_p norm, with $x_n = [\mathbf{x}]_n$ representing the n^{th} element of \mathbf{x} . We use $\mathbf{1}_N$ to denote the $N \times 1$ vector of ones. Deterministic quantities are denoted using serif typeface (e.g., $x, \mathbf{x}, \mathbf{X}$), while random quantities are denoted using sans-serif typeface (e.g., $\mathbf{x}, \mathbf{X}, \mathbf{X}$). For random variable \mathbf{x} , we write the probability density function (pdf) as $p_{\mathbf{x}}(\mathbf{x})$, the expectation as $\mathbb{E}\{\mathbf{x}\}$, and the variance as $\text{var}\{\mathbf{x}\}$. For a Gaussian random variable \mathbf{x} with mean m and variance v , we write the pdf as $\mathcal{N}(\mathbf{x}; m, v)$ and, for the special case of $\mathcal{N}(\mathbf{x}; 0, 1)$, we abbreviate the pdf as $\varphi(\mathbf{x})$ and write the complimentary cdf as $\Phi_c(\mathbf{x})$. Finally, we use $\delta(\mathbf{x})$ (where $\mathbf{x} \in \mathbb{R}$) to denote the Dirac delta distribution and δ_n (where $n \in \mathbb{Z}$) to denote the Kronecker delta sequence.

II. SIGNAL AND OBSERVATION MODELS

A. Background on BiG-AMP

As described in the introduction, a distinguishing feature of our approach is the use of BiG-AMP [30] for bilinear inference. We begin by overviewing BiG-AMP, since its operating assumptions affect the construction of our statistical model.

Consider the problem of estimating the elements of the matrices $\mathbf{S} \in \mathbb{R}^{M \times N}$ and $\mathbf{A} \in \mathbb{R}^{N \times T}$ from a noisy observation

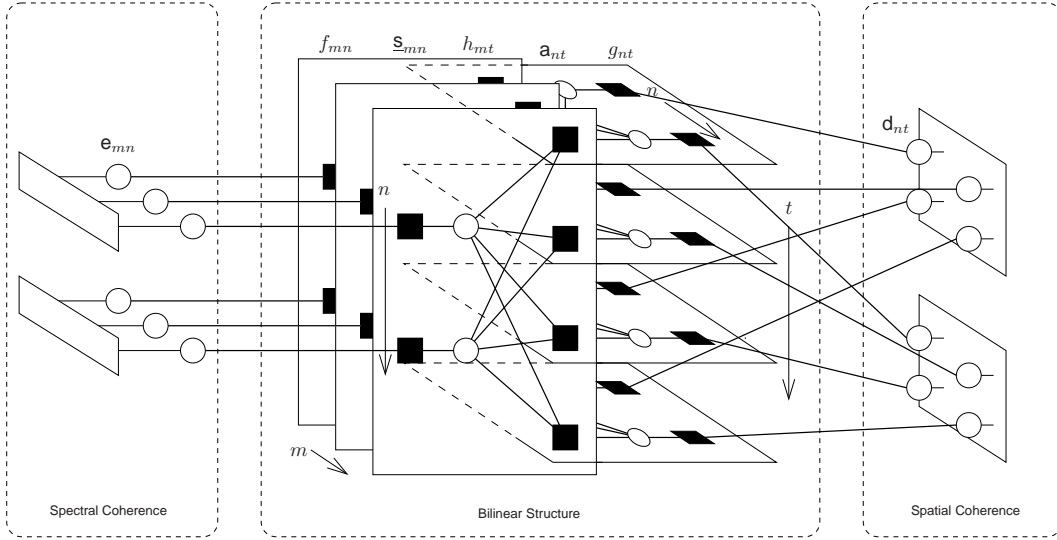


Fig. 1. The factor graph for HUT-AMP for the toy-problem dimensions $M = 3$, $N = 2$, and $T = 4$. Circles represent random variables and dark squares represent pdf factors. Each elongated bar in the left subgraph conglomerates the factors associated with an M -variable Markov chain (detailed in Fig. 2), while each square in the right subgraph conglomerates the factors associated with a $T_1 \times T_2$ -pixel Markov random field (detailed in Fig. 3).

$\bar{\mathbf{Y}} \in \mathbb{R}^{M \times T}$ of the hidden bilinear form $\bar{\mathbf{Z}} \triangleq \bar{\mathbf{S}}\bar{\mathbf{A}} \in \mathbb{R}^{M \times T}$. (Our use of overbar notation will become clear in the sequel.) Suppose that the elements of both $\bar{\mathbf{S}}$ and $\bar{\mathbf{A}}$ can be modeled as *independent* random variables \bar{s}_{mn} and \bar{a}_{nt} with known prior pdfs $p_{\bar{s}_{mn}}(\cdot)$ and $p_{\bar{a}_{nt}}(\cdot)$, respectively, with \bar{s}_{mn} being zero-mean. Suppose also that the likelihood function of $\bar{\mathbf{Z}}$ is known and *separable*, i.e., of the form

$$p_{\bar{\mathbf{Y}}|\bar{\mathbf{Z}}}(\bar{\mathbf{Y}}|\bar{\mathbf{Z}}) = \prod_{m=1}^M \prod_{t=1}^T p_{\bar{y}_{mt}|\bar{z}_{mt}}(\bar{y}_{mt}|\bar{z}_{mt}). \quad (3)$$

Finally, suppose that the dimensions M, N, T are sufficiently large. In this case, approximations of the marginal posterior pdfs of \bar{s}_{mn} , \bar{a}_{nt} , and \bar{z}_{mt} can tractably be computed using loopy belief propagation (LBP) [27], and in particular using an approximation of the sum-product algorithm (SPA) [43] known as BiG-AMP [30]. More precisely, BiG-AMP approximates the marginal posterior pdf of \bar{s}_{mn} as

$$p_{\bar{s}_{mn}|\bar{\mathbf{Y}}}(\bar{s}_{mn}|\bar{\mathbf{Y}}) = \frac{p_{\bar{s}_{mn}}(\bar{s}_{mn})\mathcal{N}(\bar{s}_{mn}; \hat{q}_{mn}, \nu_{mn}^q)}{\int p_{\bar{s}_{mn}}(\bar{s}'_{mn})\mathcal{N}(\bar{s}'_{mn}; \hat{q}_{mn}, \nu_{mn}^q)ds'_{mn}}, \quad (4)$$

where the parameters \hat{q}_{mn} and ν_{mn}^q are iteratively updated at each BiG-AMP iteration; similar approximations are made for the marginal posteriors of \bar{a}_{nt} and \bar{z}_{mt} . BiG-AMP also computes the means and variances of these approximate marginal posteriors at each iteration, yielding approximate MMSE estimates of \bar{s}_{mn} , \bar{a}_{nt} , and \bar{z}_{mt} , as well as approximations of their corresponding MSEs. For many priors of interest (e.g., the ones used in this paper), these means and variances can be computed in closed form.

In the big picture, BiG-AMP can be understood as a recent generalization of the AMP methods [31]–[33] from linear to bilinear inference. These AMP methods can be derived by starting with the SPA and applying i) central-limit-theorem arguments that approximate all messages as Gaussian and ii) Taylor-series approximations that reduce the number of

messages. Under additional independence and sub-Gaussianity assumptions, these AMP methods can be analyzed in the large-system limit, where their behavior is fully characterized by a state evolution [44]. When the state evolution has a unique fixed point, the posterior approximations computed by AMP are in fact Bayes-optimal in the large-system limit [44]. For finite-sized problems, the fixed points of AMP are known to coincide with the stationary points of a particular Bethe free energy approximation [45], [46]. For a more detailed description of how AMP methods fit into the larger family of variational Bayesian methods, we refer the reader to the recent tutorial [47]. For a detailed derivation of BiG-AMP, we refer the reader to [30].

BiG-AMP's complexity is in general dominated by ten matrix multiplies (of the form $\bar{\mathbf{S}}\bar{\mathbf{A}}$) per iteration, although simplifications can be made in the case of Gaussian $p_{\bar{y}_{mt}|\bar{z}_{mt}}(\bar{y}_{mt}|\bar{z}_{mt})$ that reduce the complexity to three matrix multiplies per iteration [30]. Furthermore, when BiG-AMP's likelihood function and priors include unknown parameters Ω , expectation-maximization (EM) methods can be used to learn them, as described in [30]. BiG-AMP was shown [48] to yield excellent performance on matrix completion, robust PCA, and dictionary learning problems, and here we show that it performs very well on the NMF and HU problems as well.

B. Augmented Observation Model

We model the elements of the m th row of the additive noise matrix \mathbf{W} in (1) as i.i.d zero-mean Gaussian with variance $\psi_m > 0$. Thus, the BiG-AMP marginal likelihoods take the form $p_{\bar{y}_m|\bar{z}_m}(\bar{y}_m|\bar{z}_m) = \mathcal{N}(\bar{y}_m; \bar{z}_m; \psi_m)$. For now we treat ψ as known, but later (in Sec. III-C) we describe how it and other model parameters can be learned from \mathbf{Y} .

Leveraging the zero-mean property of the noise, we first perform mean-removal on the observations \mathbf{Y} . In particular,

we subtract the empirical mean

$$\mu \triangleq \frac{1}{MT} \sum_{t=1}^T \sum_{m=1}^M y_{mt} = \frac{1}{MT} \mathbf{1}_M^\top \mathbf{Y} \mathbf{1}_T \quad (5)$$

from \mathbf{Y} to obtain

$$\underline{\mathbf{Y}} \triangleq \mathbf{Y} - \mu \mathbf{1}_M \mathbf{1}_T^\top \quad (6)$$

$$= \underbrace{(\mathbf{S} - \mu \mathbf{1}_M \mathbf{1}_N^\top)}_{\triangleq \underline{\mathbf{S}}} \mathbf{A} + \mathbf{W}, \quad (7)$$

where (7) employed (1) and $\mathbf{1}_N^\top \mathbf{A} = \mathbf{1}_T^\top$, the latter of which results from the simplex constraint on the columns of \mathbf{A} . It can then be shown (see Appendix A) that the elements of $\underline{\mathbf{S}}$ in (7) are approximately zero-mean.

To enforce the linear equality constraint $\mathbf{1}_N^\top \mathbf{A} = \mathbf{1}_T^\top$, we augment the observation model (7) into the form

$$\underbrace{\begin{bmatrix} \mathbf{Y} \\ \mathbf{1}_T^\top \end{bmatrix}}_{\triangleq \underline{\mathbf{Y}}} = \underbrace{\begin{bmatrix} \underline{\mathbf{S}} \\ \mathbf{1}_N^\top \end{bmatrix}}_{\triangleq \underline{\mathbf{S}}} \mathbf{A} + \underbrace{\begin{bmatrix} \mathbf{W} \\ \mathbf{0}_T^\top \end{bmatrix}}_{\triangleq \underline{\mathbf{W}}}. \quad (8)$$

For the augmented model (8), the likelihood function of $\underline{\mathbf{Z}} \triangleq \underline{\mathbf{S}} \mathbf{A}$ takes the form in (3) with

$$p_{\underline{\mathbf{Y}}|\underline{\mathbf{Z}}}(\underline{\mathbf{Y}}|\underline{\mathbf{Z}}) = \underbrace{\begin{cases} \mathcal{N}(\underline{y}_{mt}; \underline{z}_{mt}, \psi_m) & m=1, \dots, M \\ \delta(\underline{y}_{mt} - \underline{z}_{mt}) & m=M+1. \end{cases}}_{\triangleq h_{mt}(\underline{z}_{mt})} \quad (9)$$

We note that, ignoring spectral and spatial coherence, the model (8) is appropriate for the application of BiG-AMP, since the likelihood function $p_{\underline{\mathbf{Y}}|\underline{\mathbf{Z}}}(\underline{\mathbf{Y}}|\underline{\mathbf{Z}})$ is known (up to ψ) and separable, and since the elements in $\underline{\mathbf{S}}$ and \mathbf{A} can be treated as independent random variables with priors known up to a set of parameters, with those in $\underline{\mathbf{S}}$ being approximately zero-mean. In the sequel, we describe how the model (8) can be extended to capture spectral and spatial coherence. As we will see, this will be done through the introduction of additional variables that allow $\underline{\mathbf{S}}$ and \mathbf{A} to be treated as *conditionally* independent.

C. Endmember Prior

We desire a model that promotes spectral coherence in the endmembers, i.e., correlation among the (mean removed) spectral amplitudes $\{\underline{\mathbf{s}}_{mn}\}_{m=1}^M$ of each material n . However, since BiG-AMP needs $\underline{\mathbf{s}}_{mn}$ to be independent, we cannot impose correlation on these variables directly. Instead, we introduce an auxiliary sequence of correlated amplitudes $\{\mathbf{e}_{mn}\}_{m=1}^M$ such that $\underline{\mathbf{s}}_{mn}$ are independent *conditional on* \mathbf{e}_{mn} . In particular,

$$p_{\underline{\mathbf{S}}|\mathbf{E}}(\underline{\mathbf{S}}|\mathbf{E}) = \prod_{m=1}^M \prod_{n=1}^N p_{\underline{\mathbf{S}}|\mathbf{e}}(\underline{\mathbf{s}}_{mn}|\mathbf{e}_{mn}) \quad (10)$$

$$p_{\underline{\mathbf{S}}|\mathbf{e}}(\underline{\mathbf{s}}_{mn}|\mathbf{e}_{mn}) = \underbrace{\delta(\underline{\mathbf{s}}_{mn} - \mathbf{e}_{mn})}_{\triangleq f_{mn}(\mathbf{s}_{mn}, \mathbf{e}_{mn})}, \quad (11)$$

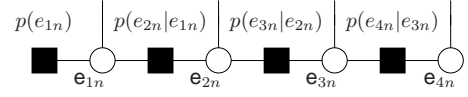


Fig. 2. Factor graph for the stationary first-order Gauss-Markov chain used to model coherence in the spectrum of the n^{th} endmember, shown here for $M = 4$ spectral bands. Incoming messages from BiG-AMP flow downward into the \mathbf{e}_{mn} nodes, and outgoing messages to BiG-AMP flow upward from the \mathbf{e}_{mn} nodes.

implying that \mathbf{e}_{mn} is merely a copy of $\underline{\mathbf{s}}_{mn}$. To impart correlation within the auxiliary sequences $\{\mathbf{e}_{mn}\}_{m=1}^M$, we model them as independent Gauss-Markov models

$$p_{\mathbf{E}}(\mathbf{E}) = \prod_{n=1}^N \underbrace{p(e_{1n}) \prod_{m=2}^M p(e_{mn}|\mathbf{e}_{m-1,n})}_{\triangleq p_{\mathbf{e}_n}(\mathbf{e}_n)}, \quad (12)$$

where $\mathbf{e}_n \triangleq [\mathbf{e}_{1n}, \dots, \mathbf{e}_{Mn}]^\top$, $\mathbf{e}_n \triangleq [e_{1n}, \dots, e_{Mn}]^\top$, and

$$p(e_{1n}) = \mathcal{N}(e_{1n}; \kappa_n, \sigma_n^2) \quad (13)$$

$$p(e_{mn}|\mathbf{e}_{m-1,n}) = \mathcal{N}(e_{mn}; (1-\eta_n)e_{m-1,n} + \eta_n \kappa_n, \eta_n^2 \sigma_n^2). \quad (14)$$

In (13)-(14), $\kappa_n \in \mathbb{R}$ controls the mean of the n^{th} process, σ_n^2 controls the variance, and $\eta_n \in [0, 1]$ controls the correlation. The resulting factor graph is illustrated in Fig. 2.

We note that the model (13)-(14) does not explicitly enforce non-negativity in \mathbf{s}_{mn} because, for simplicity, we have omitted the constraint $\underline{\mathbf{s}}_{mn} \geq -\mu$. Enforcement of $\underline{\mathbf{s}}_{mn} \geq -\mu$ could be accomplished by replacing the pdfs in (13)-(14) with truncated Gaussian versions, but the computations required for inference would become much more tedious. In our experience, this tedium is not warranted: with practical HU datasets,¹ it suffices to enforce non-negativity in \mathbf{A} and keep $\mathbf{Y} \approx \mathbf{S} \mathbf{A}$.

D. Abundance Prior

We desire a model that promotes both sparsity and spatial coherence in the abundances \mathbf{a}_{nt} . To accomplish the latter, we impose structure on the *support* of $\{\mathbf{a}_{nt}\}_{t=1}^T$ for each material n . For this purpose, we introduce the support variables $\mathbf{d}_{nt} \in \{-1, 1\}$, where $\mathbf{d}_{nt} = -1$ indicates that \mathbf{a}_{nt} is zero-valued, and $\mathbf{d}_{nt} = 1$ indicates that \mathbf{a}_{nt} is non-zero with probability 1, which we will refer to as “active.” By modeling the abundances \mathbf{a}_{nt} as independent *conditional on* \mathbf{d}_{nt} , we comply with the independence assumptions of BiG-AMP. In particular, we assume that

$$p_{\mathbf{A}|\mathbf{D}}(\mathbf{A}|\mathbf{D}) = \prod_{n=1}^N \prod_{t=1}^T p_{\mathbf{a}_n|\mathbf{d}_n}(\mathbf{a}_{nt}|\mathbf{d}_{nt}) \quad (15)$$

$$p_{\mathbf{a}_n|\mathbf{d}_n}(\mathbf{a}_{nt}|\mathbf{d}_{nt}) = \underbrace{\begin{cases} \delta(\mathbf{a}_{nt}) & \mathbf{d}_{nt} = -1 \\ \zeta_n(\mathbf{a}_{nt}) & \mathbf{d}_{nt} = 1 \end{cases}}_{\triangleq g_{nt}(\mathbf{a}_{nt}, \mathbf{d}_{nt})}, \quad (16)$$

where $\zeta_n(\cdot)$ denotes the pdf of \mathbf{a}_{nt} when active. Essentially, we employ a Bernoulli- $\zeta_n(\cdot)$ distribution for the n^{th} material.

¹Throughout our numerical experiments, the proposed inference method never produced a negative estimate of \mathbf{s}_{mn} .

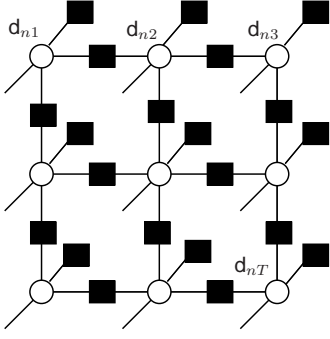


Fig. 3. Factor graph for the Ising MRF used to model spatial coherence in the support of the n th abundance map, here for $T = 3 \times 3$ pixels. Incoming messages from BiG-AMP flow diagonally upward into the d_{nt} nodes, and outgoing messages to BiG-AMP flow diagonally downward from the d_{nt} nodes.

We then place a Markov random field (MRF) prior on the support of the n th material, $\mathbf{d}_n \triangleq [d_{n1}, \dots, d_{nT}]^T$:

$$p_{\mathbf{D}}(\mathbf{D}) = \prod_{n=1}^N p_{\mathbf{d}_n}(\mathbf{d}_n) \quad (17)$$

$$p_{\mathbf{d}_n}(\mathbf{d}_n) \propto \exp \left(\sum_{t=1}^T \left(\frac{1}{2} \sum_{i \in \mathcal{D}_t} \beta_n d_{ni} - \alpha_n \right) d_{nt} \right), \quad (18)$$

where $\mathcal{D}_t \subset \{1, \dots, T\} \setminus t$ denotes the neighbors of pixel t . Roughly speaking, larger β_n yields higher spatial coherence and larger α_n yields higher sparsity. For simplicity, we adopt a neighborhood structure corresponding to the classical Ising model [28], as illustrated by the factor graph in Fig. 3.

As for the active abundances, we adopt a non-negative Gaussian mixture (NNGM) distribution for $\zeta_n(\cdot)$:

$$\zeta_n(a) = \sum_{\ell=1}^L \omega_{n\ell}^a \mathcal{N}_+(a; \theta_{n\ell}^a, \phi_{n\ell}^a), \quad (19)$$

where $\omega_{n\ell}^a \geq 0$ and $\sum_{\ell=1}^L \omega_{n\ell}^a = 1$. In (19), \mathcal{N}_+ refers to the truncated Gaussian pdf

$$\mathcal{N}_+(x; \theta, \phi) \triangleq \begin{cases} 0 & x < 0 \\ \frac{\mathcal{N}(x; \theta, \phi)}{\Phi_c(\theta/\sqrt{\phi})} & x \geq 0 \end{cases}, \quad (20)$$

where $\theta \in \mathbb{R}$ is a location parameter (but not the mean), $\phi > 0$ is a scale parameter (but not the variance), and $\Phi_c(\cdot)$ is the complimentary cdf of the $\mathcal{N}(0, 1)$ distribution. In practice, we find that $L = 3$ mixture components suffice, and we used this value throughout our numerical experiments in Sec. IV. We use a NNGM prior based on its ability to faithfully model a wide range of distributions (including multi-modal ones) and the ease by which its parameters, $\{\omega_{n\ell}^a, \theta_{n\ell}^a, \phi_{n\ell}^a\}$, can be accurately tuned using the EM method developed in [40] and discussed in Sec. III-C.

We note that the abundance model described in this section treats the abundance coefficients as correlated across pixels but statistically independent across materials. Meanwhile, the likelihood function described in Sec. II-B enforces a sum-to-one constraint across materials at each pixel. These statistical

structures are then merged in the posterior. An alternative approach that allows an abundance prior with correlation across pixels and sum-to-one across materials recently appeared in [49]. Implementing this approach in conjunction with AMP is an interesting topic for future research.

III. THE HUT-AMP ALGORITHM

A. Message Passing and Turbo Inference

Our overall goal is to jointly estimate the (correlated, non-negative) endmembers \mathbf{S} and (structured sparse, simplex-constrained) abundances \mathbf{A} from noisy observations \mathbf{Y} of the bilinear form $\mathbf{Z} = \mathbf{S}\mathbf{A}$. Using the mean-removed, augmented probabilistic models from Sec. II, the joint pdf of all random variables can be factored as follows:

$$p(\bar{\mathbf{Y}}, \bar{\mathbf{S}}, \mathbf{A}, \mathbf{E}, \mathbf{D}) = p(\bar{\mathbf{Y}}|\bar{\mathbf{S}}, \mathbf{A}) p(\bar{\mathbf{S}}, \mathbf{E}) p(\mathbf{A}, \mathbf{D}) \quad (21)$$

$$= p_{\bar{\mathbf{Y}}|\bar{\mathbf{Z}}}(\bar{\mathbf{Y}}|\bar{\mathbf{S}}\mathbf{A}) p_{\bar{\mathbf{S}}|\mathbf{E}}(\bar{\mathbf{S}}|\mathbf{E}) p_{\mathbf{E}}(\mathbf{E}) p_{\mathbf{A}|\mathbf{D}}(\mathbf{A}|\mathbf{D}) p_{\mathbf{D}}(\mathbf{D}) \quad (22)$$

$$= \left(\prod_{m=1}^{M+1} \prod_{t=1}^T h_{mt} \left(\sum_{n=1}^N \bar{s}_{mn} a_{nt} \right) \right) \times \prod_{n=1}^N \left(\delta(\bar{s}_{M+1,n} - 1) p_{\mathbf{e}_n}(\mathbf{e}_n) \prod_{m=1}^M f_{mn}(\bar{s}_{mn}, e_{mn}) \right. \\ \left. \times p_{\mathbf{d}_n}(\mathbf{d}_n) \prod_{t=1}^T g_{nt}(a_{nt}, d_{nt}) \right), \quad (23)$$

yielding the factor graph in Fig. 1. Due to the cycles within the factor graph, exact inference is NP-hard [50], and so we settle for approximate MMSE inference.

To accomplish approximate MMSE inference, we apply a form of loopy belief propagation that is inspired by the “turbo decoding” approach used in modern communications receivers [51]. In particular, after partitioning the overall factor graph into three subgraphs, as in Fig. 1, we alternate between message-passing *within* subgraphs and message-passing *between* subgraphs. In our case, BiG-AMP [30] is used for message-passing within the bilinear subgraph and standard methods from [28], [29] are used for message-passing within the other two subgraphs, which involve N Gauss-Markov chains and N binary MRFs, respectively. Overall, our proposed approach can be interpreted as a bilinear extension of the “turbo AMP” approach first proposed in [34].

B. Messaging Between Subgraphs

For a detailed description of the message passing *within* the Gauss-Markov, MRF, and BiG-AMP subgraphs, we refer interested readers to [28], [29], and [30], respectively. We now describe the message passing *between* subgraphs, which relies on the sum-product algorithm (SPA) [43]. In our implementation of the SPA, we assume that all messages are scaled to form valid pdfs (in the case of continuous random variables) or probability mass functions (pmfs) (in the case of discrete random variables), and we use $\Delta_c^b(\cdot)$ to represent the message passed from node b to node c .

As described in [43], the SPA message flowing out of a variable node along a given edge equals the (scaled) product

of messages flowing into that node along its other edges. Meanwhile, the SPA message flowing out of a factor node along a given edge equals the (scaled) integral of the product of all incoming messages times the factor associated with that node. Finally, the SPA approximates the posterior of a given random variable as the (scaled) product of messages flowing into that random variable.

As discussed in Sec. II-A, a key property of BiG-AMP is that certain messages within its sub-graph are approximated as Gaussian. In particular,

$$\Delta_{f_{mn}}^{\underline{s}_{mn}}(\underline{s}) = \mathcal{N}(\underline{s}; \hat{q}_{mn}, \nu_{mn}^q) \quad (24)$$

$$\Delta_{g_{nt}}^{\underline{a}_{nt}}(a) = \mathcal{N}(a; \hat{r}_{nt}, \nu_{nt}^r), \quad (25)$$

where the quantities $\hat{q}_{mn}, \nu_{mn}^q, \hat{r}_{nt}, \nu_{nt}^r$ are computed during the final iteration of BiG-AMP. Thus, the SPA approximated posteriors on \underline{s}_{mn} and \underline{a}_{nt} take the form

$$p_{\underline{s}_{mn}|\underline{q}_{mn}}(\underline{s}|\hat{q}_{mn}; \nu_{mn}^q) \propto \Delta_{f_{mn}}^{\underline{s}_{mn}}(\underline{s}) \mathcal{N}(\underline{s}; \hat{q}_{mn}, \nu_{mn}^q) \quad (26)$$

$$p_{\underline{a}_{nt}|\underline{r}_{nt}}(\underline{a}|\hat{r}_{nt}; \nu_{nt}^r) \propto \Delta_{g_{nt}}^{\underline{a}_{nt}}(\underline{a}) \mathcal{N}(a; \hat{r}_{nt}, \nu_{nt}^r), \quad (27)$$

where $\Delta_{f_{mn}}^{\underline{s}_{mn}}(\underline{s})$ and $\Delta_{g_{nt}}^{\underline{a}_{nt}}(a)$ can be interpreted as priors, $\mathcal{N}(\underline{s}; \hat{q}_{mn}, \nu_{mn}^q)$ and $\mathcal{N}(a; \hat{r}_{nt}, \nu_{nt}^r)$ can be interpreted as likelihoods, and (26) and (27) can be interpreted as Bayes rule. We will use these properties in the sequel.

First, we discuss the message-passing between the bilinear sub-graph and spectral-coherence sub-graph in Fig. 1. Given (11), (24), and the construction of the factor graph in Fig. 1, the SPA implies that

$$\Delta_{e_{mn}}^{f_{mn}}(e) \propto \int f_{mn}(\underline{s}, e) \Delta_{f_{mn}}^{\underline{s}_{mn}}(\underline{s}) d\underline{s} \quad (28)$$

$$= \mathcal{N}(e; \hat{q}_{mn}, \nu_{mn}^q). \quad (29)$$

The messages in (29) are used as inputs to the Gauss-Markov inference procedure. By construction, the outputs of the Gauss-Markov inference procedure will also be Gaussian beliefs. Denoting their means and variances by $\theta_{mn}^{\underline{s}_{mn}}$ and $\phi_{mn}^{\underline{s}_{mn}}$, respectively, we have that

$$\Delta_{f_{mn}}^{\underline{e}_{mn}}(e) \propto \mathcal{N}(e; \theta_{mn}^{\underline{s}_{mn}}, \phi_{mn}^{\underline{s}_{mn}}) \quad (30)$$

$$\Delta_{\underline{s}_{mn}}^{f_{mn}}(\underline{s}) = \int f_{mn}(\underline{s}, e) \Delta_{f_{mn}}^{\underline{e}_{mn}}(e) de \quad (31)$$

$$= \mathcal{N}(\underline{s}; \theta_{mn}^{\underline{s}_{mn}}, \phi_{mn}^{\underline{s}_{mn}}). \quad (32)$$

When BiG-AMP is subsequently called for inference on the bilinear sub-graph, (32) is inserted into (26), i.e., $\Delta_{\underline{s}_{mn}}^{f_{mn}}(\cdot)$ acts as the prior on \underline{s}_{mn} .

Next we discuss the message-passing between the bilinear sub-graph and the spatial-coherence sub-graph in Fig. 1. The SPA, together with the construction of the factor graph in Fig. 1, imply

$$\Delta_{\underline{a}_{nt}}^{g_{nt}}(d) = \frac{\int g_{nt}(a, d) \Delta_{g_{nt}}^{\underline{a}_{nt}}(a) da}{\sum_{d'=\pm 1} \int g_{nt}(a, d') \Delta_{g_{nt}}^{\underline{a}_{nt}}(a) da}, \quad d \in \pm 1. \quad (33)$$

Given (16) and (25), we find that

$$\begin{aligned} & \int g_{nt}(a, d) \Delta_{g_{nt}}^{\underline{a}_{nt}}(a) da \\ &= \begin{cases} \mathcal{N}(0; \hat{r}_{nt}, \nu_{nt}^r) da & d = -1 \\ \int \zeta_n(a) \mathcal{N}(a; \hat{r}_{nt}, \nu_{nt}^r) da & d = 1 \end{cases} \quad (34) \end{aligned}$$

which implies

$$\Delta_{\underline{a}_{nt}}^{g_{nt}}(d = +1) = \left(1 + \frac{\mathcal{N}(0; \hat{r}_{nt}, \nu_{nt}^r)}{\int \zeta_n(a) \mathcal{N}(a; \hat{r}_{nt}, \nu_{nt}^r) da} \right)^{-1} \quad (35a)$$

$$\Delta_{\underline{a}_{nt}}^{g_{nt}}(d = -1) = 1 - \Delta_{\underline{a}_{nt}}^{g_{nt}}(d = +1), \quad (35b)$$

where the fraction in (35a) is BiG-AMP's approximation of the likelihood ratio $p_{\mathbf{Y}|\mathbf{d}_{nt}}(\mathbf{Y}|-1)/p_{\mathbf{Y}|\mathbf{d}_{nt}}(\mathbf{Y}|+1)$.

The Bernoulli beliefs from (35) are used as inputs to the MRF-based support-inference procedure. The outputs of the MRF inference procedure will also be Bernoulli beliefs of the form

$$\Delta_{g_{nt}}^{\underline{d}_{nt}}(d = +1) = \pi_{nt} \quad (36a)$$

$$\Delta_{g_{nt}}^{\underline{d}_{nt}}(d = -1) = 1 - \pi_{nt} \quad (36b)$$

for some $\pi_{nt} \in (0, 1)$. The SPA and (16) then imply that

$$\Delta_{\underline{a}_{nt}}^{g_{nt}}(a) \propto \sum_{d=\pm 1} g_{nt}(a, d) \Delta_{g_{nt}}^{\underline{d}_{nt}}(d) \quad (37)$$

$$= (1 - \pi_{nt})\delta(a) + \pi_{nt}\zeta_n(a) \quad (38)$$

for $\zeta_n(\cdot)$ defined in (19). When BiG-AMP is subsequently called for inference on the bilinear sub-graph, (38) is inserted into (27), i.e., $\Delta_{\underline{a}_{nt}}^{g_{nt}}(\cdot)$ acts as the prior on \underline{a}_{nt} .

C. EM Learning of the Prior Parameters

In practice, we desire that the parameters

$$\Omega = \left\{ \psi, \{\omega_{n\ell}^a, \theta_{n\ell}^a, \phi_{n\ell}^a\}_{\forall n\ell}, \{\eta_n, \kappa_n, \sigma_n^2, \alpha_n, \beta_n\}_{\forall n} \right\} \quad (39)$$

used for the assumed likelihood $p_{\mathbf{y}_{mt}|\mathbf{z}_{mt}}(\mathbf{y}_{mt}|\cdot)$, NNGM abundance prior $\zeta_n(\cdot)$, Gauss-Markov chain $p_{\mathbf{e}_n}(\cdot)$, and binary MRF $p_{\mathbf{d}_n}(\cdot)$ are well tuned. With this in mind, we propose an expectation-maximization (EM) [52] procedure to tune Ω , similar to that used for the GAMP-based sparse-reconstruction algorithms in [53] and [40].

To tune Ω , the EM algorithm [52] iterates

$$\Omega^{i+1} = \arg \max_{\Omega} \mathbb{E} \{ \ln p(\mathbf{E}, \mathbf{A}, \mathbf{D}, \bar{\mathbf{Y}}; \Omega) | \bar{\mathbf{Y}}; \Omega^i \} \quad (40)$$

with the goal of increasing a lower bound on the true likelihood $p(\bar{\mathbf{Y}}; \Omega)$ at each EM-iteration i . In our case, the true posterior distribution used to evaluate the expectation in (40) is NP-hard to compute, and so we use the SPA-approximated posteriors $\widehat{p}_{\mathbf{E}|\bar{\mathbf{Y}}}(\mathbf{E}|\bar{\mathbf{Y}}) \propto \prod_{m,n} \Delta_{e_{mn}}^{f_{mn}}(e_{mn}) \Delta_{f_{mn}}^{\underline{e}_{mn}}(e_{mn})$ from (29)-(30), $\widehat{p}_{\mathbf{D}|\bar{\mathbf{Y}}}(\mathbf{D}|\bar{\mathbf{Y}}) \propto \prod_{n,t} \Delta_{\underline{d}_{nt}}^{g_{nt}}(d_{nt}) \Delta_{g_{nt}}^{\underline{d}_{nt}}(d_{nt})$ from (35)-(36), and $\widehat{p}_{\mathbf{A}|\bar{\mathbf{Y}}}(\mathbf{A}|\bar{\mathbf{Y}}) \propto \prod_{n,t} \Delta_{\underline{a}_{nt}}^{g_{nt}}(a_{nt}) \Delta_{g_{nt}}^{\underline{a}_{nt}}(a_{nt})$ from (25) and (38). Furthermore, since it is difficult to perform the maximization in (40) jointly, we maximize Ω one component at a time (while holding the others fixed), which is the well known ‘‘incremental’’ variant of EM [54].

The resulting EM-update expressions for the noise and NNGM parameters $\psi, \omega_{n\ell}^a, \theta_{n\ell}^a, \phi_{n\ell}^a$ can be found in [40], and those for the Gauss-Markov chain parameters $\eta_n, \kappa_n, \sigma_n^2$ can be found in [39]. They are all computed in closed-form using readily available quantities, and thus do not add significantly to the complexity of HUT-AMP. The update procedure for the binary MRF parameters α_n, β_n is described in [38] and uses gradient descent. Since a small number of gradient-descent iterations suffice, this latter procedure does not significantly increase the complexity of HUT-AMP.

D. EM Initialization

Since the EM algorithm may converge to a local maximum of the likelihood, care must be taken when initializing the EM-learned parameters. Below, we propose an initialization strategy for HUT-AMP that, based on our empirical experience, seems to work well.

We first initialize the endmembers \mathbf{S} . For this, we found it effective to use an off-the-shelf EE algorithm like VCA [7] or FSNMF² [10] to recover $\hat{\mathbf{S}}^0$. Then, as described in (7), we subtract the observation mean μ from $\hat{\mathbf{S}}^0$ to obtain the initialization $\underline{\mathbf{S}}^0$.

With the aid of $\underline{\mathbf{S}}^0$, we next run BiG-AMP under

- 1) the trivial endmember prior

$$\Delta_{\underline{\mathbf{S}}_{mn}}^{f_{mn}}(\underline{s}) = \delta(\underline{s} - \hat{\mathbf{S}}_{mn}^0), \quad (41)$$

which essentially fixes the endmembers at $\hat{\mathbf{S}}^0$,

- 2) the agnostic NNGM abundance initialization from [40]:

$$\Delta_{\mathbf{a}_{nt}}^{g_{nt}}(a) = (1 - \pi_{nt}^0)\delta(a) + \pi_{nt}^0 \sum_{\ell=1}^L \omega_{n\ell}^a \mathcal{N}_+(a; \theta_{n\ell}^a, \phi_{n\ell}^a) \quad (42)$$

with $\{\omega_{n\ell}^a, \theta_{n\ell}^a, \phi_{n\ell}^a\}_{\ell=1}^L$ set at the best fit to a uniform distribution on the interval $[0, 1]$ and $\pi_{nt}^0 = \frac{1}{2}$, and

- 3) the noise variance initialization from [40]:

$$\psi_m^0 = \frac{\|\mathbf{Y}\|_F^2}{(\text{SNR}_m^0 + 1)MT} \quad \forall m, \quad (43)$$

where, without any prior knowledge of the true $\text{SNR}_m \triangleq \mathbb{E}\{|\mathbf{z}_{mt}|^2\}/\psi_m$, we suggest $\text{SNR}_m^0 = 10$ dB.

By running BiG-AMP under these settings, we initialize the messages $\Delta_{\underline{\mathbf{f}}_{mn}}^{\mathbf{S}_{mn}}(\cdot)$ and $\Delta_{\underline{\mathbf{g}}_{nt}}^{\mathbf{a}_{nt}}(\cdot)$ from (24)-(25) and we also obtain an initial estimate of \mathbf{A} from the mean of the approximate posterior (27), which we shall refer to as $\hat{\mathbf{A}}^0$.

Finally, we initialize the remaining parameters in Ω . Starting with the spectral coherence parameters, we set the mean κ_n^0 and variance $(\sigma_n^2)^0$ at the empirical mean and variance, respectively, of the elements in the n th column of $\hat{\mathbf{S}}^0$. Then, we initialize the correlation η_n as suggested in [39], i.e.,

$$\varphi_m^0 = \frac{\|\mathbf{y}_m\|_2^2 - T\psi_m^0}{\|\hat{\mathbf{A}}^0\|_F^2} \quad (44)$$

$$\eta_n^0 = 1 - \frac{1}{M-1} \sum_{m=1}^{M-1} \frac{|\mathbf{y}_m^\top \mathbf{y}_{m+1}|}{\varphi_m^0 \|\hat{\mathbf{A}}^0\|_F^2} \quad \text{for } n = 1, \dots, N, \quad (45)$$

where \mathbf{y}_m^\top denotes the m th row of \mathbf{Y} . Lastly, we initialize the spatial coherence parameters as suggested in [38], i.e., $\beta_n^0 = 0.4$ and $\alpha_n^0 = 0.4$, since [38] shows these values to work well over a wide operating range.

² With FSNMF (which was used for all of the experiments in Sec. IV), we found that it helped to post-process the observations to reduce the effects of noise. For this, we used the standard PCA-based denoising approach described in [3]: the signal subspace was estimated from the left singular vectors of \mathbf{Y} after row-wise mean-removal, and the FSNMF-estimated endmembers were projected onto the signal subspace.

Definitions:

$$\begin{aligned} \Delta_{\underline{\mathbf{F}}}^{\mathbf{E}} &\triangleq \{\Delta_{\underline{\mathbf{f}}_{mn}}^{\mathbf{e}_{mn}}(\cdot)\}_{\forall mn} & \Delta_{\underline{\mathbf{E}}}^{\mathbf{F}} &\triangleq \{\Delta_{\underline{\mathbf{e}}_{mn}}^{\mathbf{f}_{mn}}(\cdot)\}_{\forall mn} \\ \Delta_{\underline{\mathbf{S}}}^{\mathbf{F}} &\triangleq \{\Delta_{\underline{\mathbf{s}}_{mn}}^{\mathbf{f}_{mn}}(\cdot)\}_{\forall mn} & \Delta_{\underline{\mathbf{F}}}^{\mathbf{S}} &\triangleq \{\Delta_{\underline{\mathbf{s}}_{mn}}^{\mathbf{f}_{mn}}(\cdot)\}_{\forall mn} \\ \Delta_{\underline{\mathbf{A}}}^{\mathbf{G}} &\triangleq \{\Delta_{\underline{\mathbf{g}}_{nt}}^{\mathbf{a}_{nt}}(\cdot)\}_{\forall nt} & \Delta_{\underline{\mathbf{G}}}^{\mathbf{A}} &\triangleq \{\Delta_{\underline{\mathbf{a}}_{nt}}^{\mathbf{g}_{nt}}(\cdot)\}_{\forall nt} \\ \Delta_{\underline{\mathbf{D}}}^{\mathbf{G}} &\triangleq \{\Delta_{\underline{\mathbf{d}}_{nt}}^{\mathbf{g}_{nt}}(\cdot)\}_{\forall nt} & \Delta_{\underline{\mathbf{G}}}^{\mathbf{D}} &\triangleq \{\Delta_{\underline{\mathbf{d}}_{nt}}^{\mathbf{g}_{nt}}(\cdot)\}_{\forall nt} \end{aligned}$$

```

1: Initialize  $\Delta_{\underline{\mathbf{F}}}^{\mathbf{S}}, \Delta_{\underline{\mathbf{G}}}^{\mathbf{A}}$ , and  $\Omega^0$  as described in Sec. III-D.
2: for  $i = 1, 2, 3, \dots$  do
3:   convert  $\Delta_{\underline{\mathbf{F}}}^{\mathbf{S}}$  to  $\Delta_{\underline{\mathbf{E}}}^{\mathbf{F}}$  via (24) and (29)
4:   convert  $\Delta_{\underline{\mathbf{G}}}^{\mathbf{A}}$  to  $\Delta_{\underline{\mathbf{D}}}^{\mathbf{G}}$  via (25) and (35)
5:    $\Delta_{\underline{\mathbf{E}}}^{\mathbf{E}} = \text{GaussMarkov}(\Delta_{\underline{\mathbf{E}}}^{\mathbf{F}}, \Omega^i)$ 
6:    $\Delta_{\underline{\mathbf{D}}}^{\mathbf{D}} = \text{MRF}(\Delta_{\underline{\mathbf{D}}}^{\mathbf{G}}, \Omega^i)$ 
7:   convert  $\Delta_{\underline{\mathbf{E}}}^{\mathbf{E}}$  to  $\Delta_{\underline{\mathbf{S}}}^{\mathbf{F}}$  via (30) and (32)
8:   convert  $\Delta_{\underline{\mathbf{D}}}^{\mathbf{D}}$  to  $\Delta_{\underline{\mathbf{A}}}^{\mathbf{G}}$  via (36) and (38)
9:    $\Omega^i = \text{EM}(\Delta_{\underline{\mathbf{E}}}^{\mathbf{E}}, \Delta_{\underline{\mathbf{F}}}^{\mathbf{S}}, \Delta_{\underline{\mathbf{D}}}^{\mathbf{G}}, \Delta_{\underline{\mathbf{A}}}^{\mathbf{G}}, \Delta_{\underline{\mathbf{G}}}^{\mathbf{A}}, \Omega^{i-1})$ 
10:   $[\Delta_{\underline{\mathbf{F}}}^{\mathbf{S}}, \Delta_{\underline{\mathbf{G}}}^{\mathbf{A}}] = \text{BiGAMP}(\Delta_{\underline{\mathbf{S}}}^{\mathbf{F}}, \Delta_{\underline{\mathbf{A}}}^{\mathbf{G}}, \Omega^i)$ 
11: end for

```

TABLE I
HUT-AMP PSEUDOCODE FOR FIXED NUMBER OF MATERIALS N .

E. HUT-AMP Summary

We now describe the scheduling of turbo-messaging and EM-tuning steps, which together constitute the HUT-AMP algorithm. Essentially, we elect to perform one EM update per turbo iteration, yielding the steps tabulated in Table I. As previously mentioned, the “**BiGAMP**” operation iterates the BiG-AMP algorithm to convergence as described in [30], the “**GaussMarkov**” operation performs standard Gauss-Markov inference as described in [28], and the “**MRF**” operation performs MRF inference via the belief-propagation method described in [29].

F. Selection of Model Order N

In practice, the number of materials N present in a scene may be unknown. Previous approaches such as the hyperspectral signal subspace identification by minimum error (HySime) [55], and a Neyman-Pearson detection theory-based thresholding method (HFC) [56] directly address the problem of estimating the number of materials N .

As an alternative, we apply a standard penalized log-likelihood maximization [42] method to estimate N from the observed data \mathbf{Y} . Specifically, we aim to solve

$$\hat{N} = \arg \max_N 2 \ln p_{\mathbf{Y}|\mathbf{Z}}(\mathbf{Y}|\hat{\mathbf{S}}_N \hat{\mathbf{A}}_N; \hat{\psi}_{\text{ML}}) - \gamma(N), \quad (46)$$

where $\hat{\mathbf{S}}_N$ and $\hat{\mathbf{A}}_N$ are the estimates of the mean-removed endmembers and abundances returned from N -material HUT-AMP, $\hat{\psi}_{\text{ML}}$ is the ML estimate of the noise variance, and $\gamma(\cdot)$ is a penalty term. As recommended in [48], we choose $\gamma(\cdot)$ in accordance with the small-sample-corrected Akaike information criterion (AICc) [42], i.e., $\gamma(N) = 2 \frac{MT}{MT - n(N) - 1} n(N)$, where MT is the number of scalar observations in \mathbf{Y} and $n(N)$ is the number of scalar degrees-of-freedom (DoF) in our model, which depends on N . In particular, $n(N)$ comprises MN DoF from \mathbf{S} , $(N-1)T$ DoF from \mathbf{A} , and $5N + 2NL + N(L-1) + M$ DoF from Ω . Plugging the standard form of the ML estimate of ψ (see, e.g., [42, eq.

(7)] into (46), we obtain

$$\hat{N} = \arg \max_N -MT \ln \left(\frac{\|\mathbf{Y} - \hat{\mathbf{S}}_N \hat{\mathbf{A}}_N\|_F^2}{MT} \right) - \frac{2MTn(N)}{MT - n(N) - 1}. \quad (47)$$

To solve the maximization in (47), we first run $N = 2$ HUT-AMP to completion and compute the penalized log-likelihood. We then increment N by 1, and compute the penalized log-likelihood again. If it increases, we increment N by 1 and repeat the procedure. Once the penalized log-likelihood decreases, we stop the procedure and select the previous model order N , which is the local maximizer of the penalized log-likelihood. We refer to the resulting procedure as “HUT-AMP with model-order selection” (HUT-AMP-MOS).

We also note that a similar model-order selection strategy can be implemented to tune the number of NNGM components L used in (19), and we refer interested readers to [53] for more details. We note, however, that the fixed choice $L = 3$ was sufficient to yield the excellent results in Sec. IV.

IV. NUMERICAL RESULTS

In this section, we report the results of several experiments that we conducted to characterize the performance of our proposed methods on both synthetic and real-world datasets.

In these experiments, we compared the endmembers $\hat{\mathbf{S}}$ recovered from our proposed HUT-AMP and HUT-AMP-MOS³ unmixing algorithms to those recovered by the Bayesian unmixing algorithm SCU [26]; the sparse NMF techniques $L_{\frac{1}{2}}$ NMF [18] and SDSNMF [19]; and the endmember extraction (EE) algorithms VCA [7], FSNMF [10], and MVSA [11].

We also compared the abundances $\hat{\mathbf{A}}$ recovered by our proposed HUT-AMP and HUT-AMP-MOS unmixing algorithms to those recovered by SCU and SDSNMF, as well as those recovered by FCLS (2) (implemented via Matlab’s `lsqlin`) and SUnSAL-TV [16] using the endmember estimates produced by VCA, FSNMF, and MVSA. We note that SCU, SDSNMF, and SUnSAL-TV all exploit spatial coherence, and that SDSNMF is in fact $L_{\frac{1}{2}}$ NMF with additional mechanisms to exploit spatial coherence.

In all cases, algorithms were run using their authors’ implementation and suggested default settings, unless noted otherwise. The only exceptions are SDSNMF and $L_{\frac{1}{2}}$ NMF, which we implemented in MATLAB since their authors declined to provide source code. All algorithms (with the exception of HUT-AMP-MOS) were supplied the true number of materials N in each experiment. For SUnSAL-TV, the regularization weights for the ℓ_1 and TV norms were hand-tuned, because cross-validation tuning was too computationally expensive given the sizes of the datasets. For FSNMF, we used the PCA post-processing described in footnote 2 to reduce the effects of measurement noise, since this greatly improved its mean-squared estimation error.

A. Pixel Purity versus Abundance Sparsity

Our first experiment aims to assess EE performance as a function of pixel purity and abundance sparsity. Our motivation stems from the fact that the proposed HUT-AMP

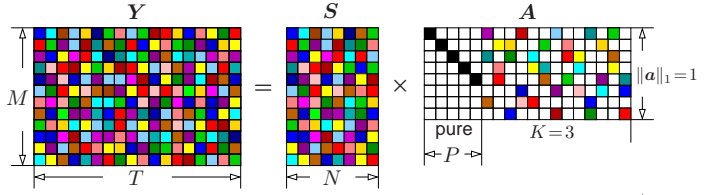


Fig. 4. Illustration of the non-negative endmember matrix \mathbf{S} and the K -sparse P -pure abundance matrix \mathbf{A} for the first experiment.

algorithm aims to exploit *sparsity* in the columns of the abundance matrix \mathbf{A} , while classical EE techniques like VCA and FSNMF aim to exploit the presence of *pure pixels*, recalling the discussion in Sec. I. Thus, we are interested in seeing how these contrasting approaches fare under varying combinations of pixel purity and abundance sparsity. We also compare against the minimum-volume-simplex approach from [11], which is an alternative to both pixel purity and abundance sparsity.

We first constructed synthetic data consisting of $M = 100$ spectral bands, $T = 115$ spatial pixels, and $N = 10$ materials. The endmember matrix $\mathbf{S} \in \mathbb{R}_+^{M \times N}$ was drawn i.i.d such that $\mathbf{S}_{mn} \sim \mathcal{N}_+(0.5, 0.05)$. The abundance matrix $\mathbf{A} \in \mathbb{R}_+^{N \times T}$ was generated as shown in Fig. 4, where P of the columns of \mathbf{A} were assigned (uniformly at random) to be pure pixels, and the remaining columns were drawn K -sparse on the simplex. In particular, for each of these latter columns, the support was drawn uniformly at random, and the non-zero values $\{\underline{a}_k\}_{k=1}^K$ were drawn from a Dirichlet distribution, i.e.,

$$p(\underline{a}_1, \dots, \underline{a}_{K-1}) = \begin{cases} \frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} \prod_{k=1}^K \underline{a}_k^{\alpha-1}, & \underline{a}_k \in [0, 1] \\ 0 & \text{else} \end{cases} \quad (48a)$$

$$p(\underline{a}_K | \underline{a}_1, \dots, \underline{a}_{K-1}) = \delta(1 - \underline{a}_1 - \dots - \underline{a}_K), \quad (48b)$$

where $\Gamma(\cdot)$ denotes the gamma function, with concentration parameter $\alpha = 1$. Finally, the observation matrix \mathbf{Y} was created by adding white Gaussian noise \mathbf{W} to $\mathbf{Z} = \mathbf{S}\mathbf{A}$, where the noise variance ψ was adjusted to achieve $\text{SNR} \triangleq \frac{1}{MT} \|\mathbf{Z}\|_F^2 / \psi = 80$ dB.

Figure 5 shows empirical success probability averaged over $R = 100$ realizations, as a function of pixel purity P and sparsity K , for the HUT-AMP, MVSA, VCA, and $L_{\frac{1}{2}}$ NMF algorithms.⁴ It does not show FSNMF since its performance was indistinguishable from VCA’s performance. Here, a recovery was considered successful if $\text{NMSE}_S \triangleq \|\mathbf{S} - \hat{\mathbf{S}}\|_F^2 / \|\mathbf{S}\|_F^2 < -40$ dB. As seen in Fig. 5(c) and Fig. 5(d), VCA and FSNMF were only successful for the $K=1$ and $P=10$ cases, i.e., the pure-pixel cases. $L_{\frac{1}{2}}$ NMF did slightly better, with successful recovery for $K \leq 2$. HUT-AMP, on the other hand, was able to successfully recover the endmembers for $K \leq 6$ -sparse abundances, even when there was only $P = 1$ pure-pixels available. We attribute HUT-AMP’s improved performance to its exploitation of sparsity rather than pure pixels (as with

⁴Since our experimental findings into sparsity-versus-purity would be biased if the algorithms under test used different approaches to the exploitation of spatial and/or spectral coherence, we turn off the coherence-exploiting mechanisms in HUT-AMP and SDSNMF (reducing the latter to $L_{\frac{1}{2}}$ NMF) and compare to other algorithms that do not exploit spatial or spectral coherence: VCA and MVSA.

³Matlab code can be found at <http://www.ece.osu.edu/~schniter/HUTAMP>.

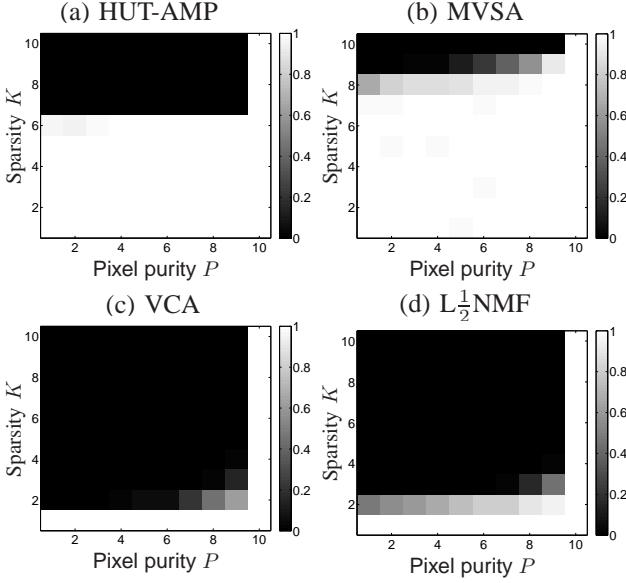


Fig. 5. First experiment: Average success rate for near-perfect recovery of i.i.d. endmembers \mathbf{S} and K -sparse and P -pure abundances \mathbf{A} using (a) HUT-AMP, (b) MVSA, (c) VCA, and (d) $L_{\frac{1}{2}}\text{NMF}$.

VCA and FSNMF), and its ability to accurately learn the underlying sparsity rate. Also, we conjecture that sparsity (i.e., $K > 1$ and $P < N$) is more important in practice, since the spatial resolution of the hyperspectral sensors may not guarantee pixel-purity for all materials, while sparse abundances (i.e., $K \ll N$) are more likely to hold. Finally, we note that, although MVSA performed remarkably well in this experiment, it performed relatively poorly for the experiments in Sec. IV-B through Sec. IV-D.

Next, we repeat the previous experiment at $\text{SNR} = 60$ dB with \mathbf{S} randomly selected from the USGS Digital Spectral Library splib06a,⁵ which contains laboratory-measured reflectance values for various materials over $M = 224$ spectral bands. In particular, for each Monte-Carlo realization we randomly select $N = 10$ endmembers from the library such that $\min_{i \neq j} \text{SAD}(\mathbf{s}_i, \mathbf{s}_j) \geq 15$ degrees, for *spectral angle distance*

$$\text{SAD}(\mathbf{s}_i, \mathbf{s}_j) \triangleq \arccos\left(\frac{\mathbf{s}_i^T \mathbf{s}_j}{\|\mathbf{s}_i\|_2 \|\mathbf{s}_j\|_2}\right) \quad (49)$$

Figure 6 shows the empirical success probability for the HUT-AMP, MVSA, VCA, and $L_{\frac{1}{2}}\text{NMF}$ algorithms. Although HUT-AMP's performance with USGS endmembers is not as good as with i.i.d. endmembers, it still outperformed VCA and $L_{\frac{1}{2}}\text{NMF}$. As before, MVSA has the best performance.

Finally, we perform a variation on the previous experiment that again uses randomly selected USGS endmembers. But rather than using pure and/or K -sparse abundance vectors, it uses fully mixed abundances whose N coefficients were generated from a Dirichlet distribution with concentration parameter α (recall (48)). Recall that larger values of α correspond to more dense mixing. Figure 7 reports the average $\text{NMSE}_{\mathbf{S}}$ of HUT-AMP, MVSA, VCA, and $L_{\frac{1}{2}}\text{NMF}$ versus both the

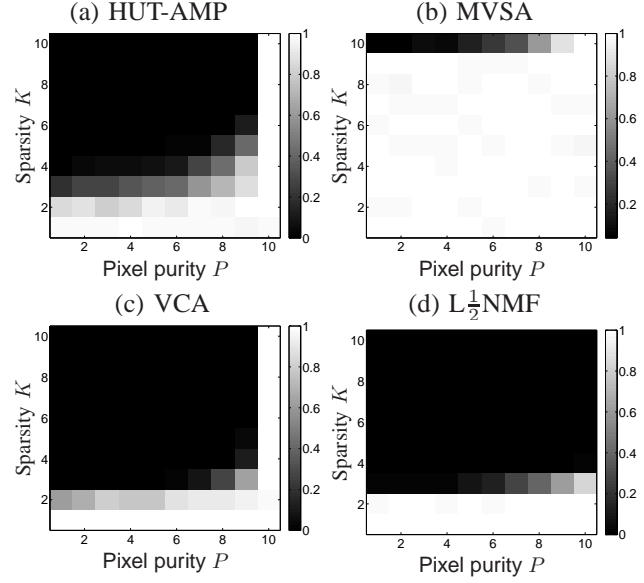


Fig. 6. Second experiment: Average success rate for near-perfect recovery of USGS endmembers \mathbf{S} and K -sparse and P -pure abundances \mathbf{A} using (a) HUT-AMP, (b) MVSA, (c) VCA, and (d) $L_{\frac{1}{2}}\text{NMF}$.

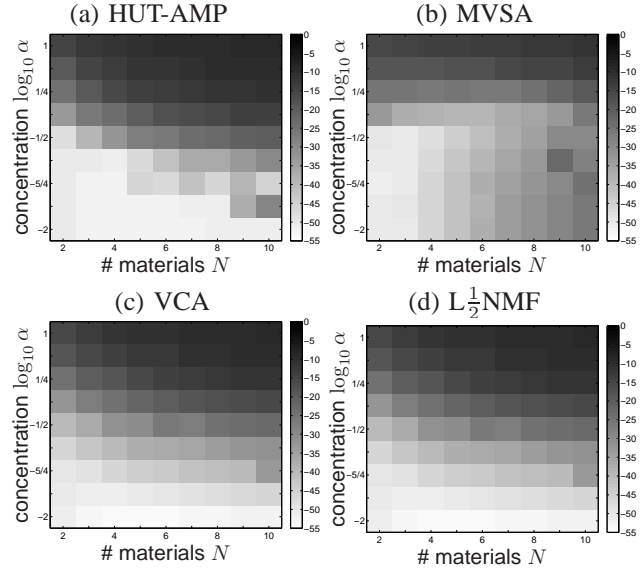


Fig. 7. Third experiment: Average $\text{NMSE}_{\mathbf{S}}$ for recovery of N USGS endmembers \mathbf{S} with abundances \mathbf{A} drawn from Dirichlet distribution with concentration α using (a) HUT-AMP, (b) MVSA, (c) VCA, and (d) $L_{\frac{1}{2}}\text{NMF}$.

number of materials, N , and the concentration parameter, α , at $\text{SNR} = 40$ dB. The figure shows that HUT-AMP, VCA, and $L_{\frac{1}{2}}\text{NMF}$ gave similar performance overall, with small advantages to HUT-AMP when $N \leq 8$ and $\alpha \leq 10^{-13/8}$. Relative to the other algorithms, MVSA tolerated higher values of α , but was more sensitive to larger numbers of materials, N , when α was small.

B. Pure-Pixel Synthetic Abundances

The second experiment uses synthetic pure-pixel abundances \mathbf{A} with endmembers \mathbf{S} chosen from the USGS Digital Spectral Library. To construct the data, we partitioned a

⁵See <http://speclab.cr.usgs.gov/spectral.lib06/ds231/>



Fig. 8. False-color image of the noiseless measurements \mathbf{Z} used for the second experiment. Since the pixels are pure, each strip shows the false color of one of the $N=5$ materials. They are, in order from left to right: Grossular, Alunite, wxl Kaolinite, Hydroxyl-Apatite, and Amphibole.

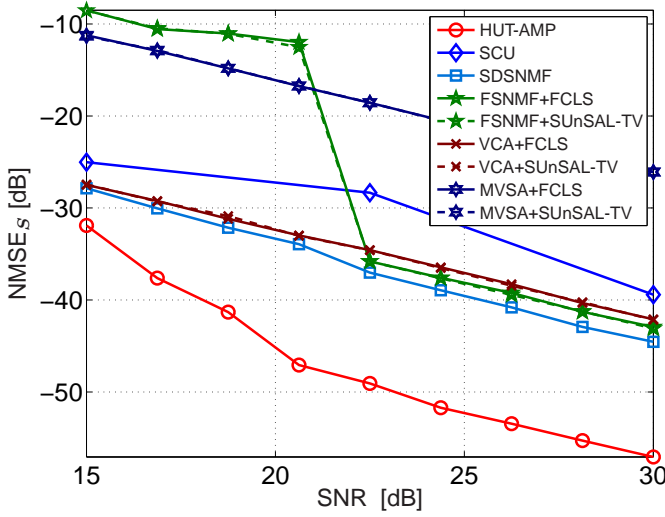


Fig. 9. NMSE_S vs. SNR for the synthetic pure pixel dataset.

scene of $T = 50 \times 50$ pixels into $N = 5$ equally sized vertical strips, each containing a single pure material. We then selected endmembers corresponding to the materials Grossular, Alunite, well crystallized (wxl) Kaolinite, Hydroxyl-Apatite, and Amphibole, noting that similar results were obtained in experiments we conducted with other materials. Figure 8 shows a false-color image constructed from the noiseless measurements \mathbf{Z} . We then vary the SNR on a grid from 15 dB to 35 dB by adding white Gaussian noise.

Averaging over $r = 50$ realizations, Figures 9 and 10 show the normalized mean-squared error of the estimated endmembers and abundances, respectively (i.e., NMSE_S and $\text{NMSE}_A \triangleq \|\mathbf{A} - \hat{\mathbf{A}}\|_F^2 / \|\mathbf{A}\|_F^2$), while Fig. 11 shows the average runtime versus SNR. For this pure-pixel dataset, these figures show HUT-AMP dominating the other algorithms in both endmember and abundance estimation accuracy at all SNRs. In particular, HUT-AMP outperformed the best competing techniques by 4 to 12 dB in NMSE_S and as much as 90 dB in NMSE_A . We note that the biggest gains in NMSE_A occurred when $\text{SNR} \geq 24$ dB.

We attribute HUT-AMP's excellent NMSE to several factors. First, it has the ability to *jointly* estimate endmembers and abundances, to exploit spectral coherence in the endmembers, and to exploit both spatial coherence and sparsity in the

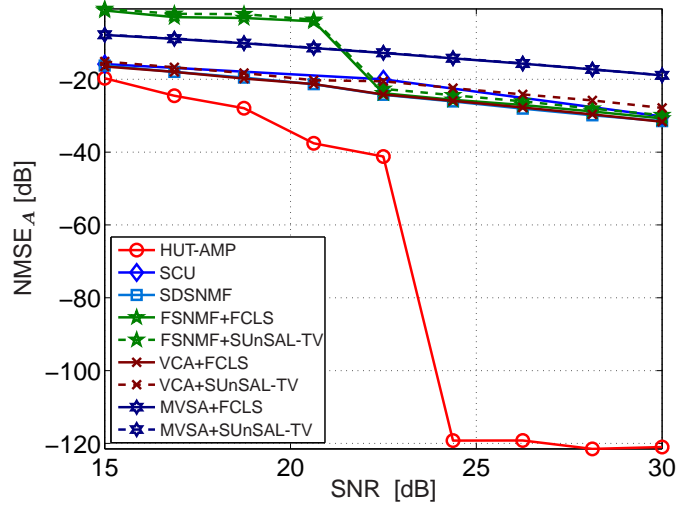


Fig. 10. NMSE_A vs. SNR for the synthetic pure pixel dataset.

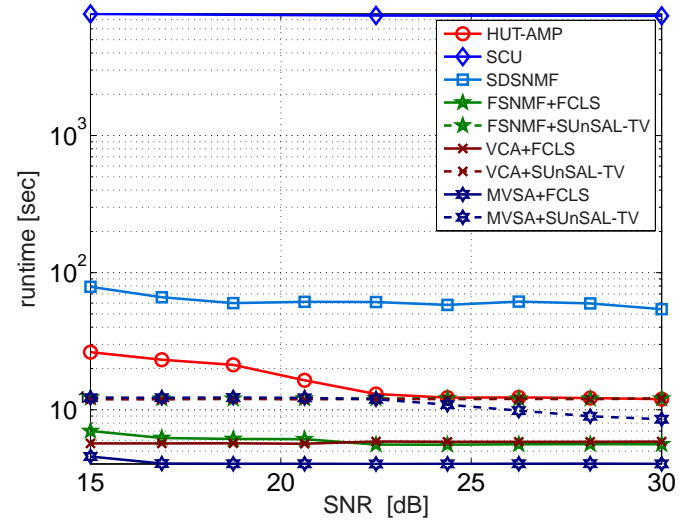


Fig. 11. Runtime vs. SNR for the synthetic pure pixel dataset.

abundances (of which there is plenty in this experiment). Furthermore, due to the presence of pure-pixels throughout the scene, the “active” distribution $\zeta_n(\cdot)$ in (19) is simply a Bernoulli distribution, which HUT-AMP is able to learn (via EM) and exploit (via BiG-AMP) for improved performance.

Figure 11 shows that the runtime of HUT-AMP was approximately 3 times slower than the EE-and-inversion techniques, approximately 4 times faster than the spatial-coherence exploiting SDSNMF algorithm, and approximately 200 times faster than that of the Bayesian SCU algorithm. We conjecture that the relatively slow runtime of SCU is due to its use of Gibbs sampling.

Although not shown in the above figures, we also ran HUT-AMP-MOS on this dataset at $\text{SNR} = 30$ dB. The result was that HUT-AMP-MOS correctly estimated the number of materials (i.e., $N = 5$) on every realization and thus gave identical NMSE_S and NMSE_A as HUT-AMP. The total runtime of HUT-AMP-MOS at this SNR was 94.27 seconds, which was about 9 times slower than HUT-AMP but still 75

times faster than SCU.

C. SHARE 2012 Avon Dataset

Next, we evaluated algorithm performance on the SHARE 2012 Avon dataset⁶ [57], which uses $M = 360$ spectral bands, corresponding to wavelengths between 400 and 2450 nm, over a large rural area. To do this, we first cropped down the full image to the scene shown in Fig. 12, which is known to consist of $N = 4$ materials: grass, dry sand, black felt, and white TyVek [58]. This scene was explicitly constructed for use in hyperspectral unmixing experiments, as efforts were made to ensure that the vast majority of the pixels were pure. Also, the data was collected on a nearly cloudless day, implying that shadowing effects were minimal. To construct ground-truth endmembers,⁷ we averaged a 4×4 pixel grid of the received spectra in a “pure” region for each material. We then computed SAD between each ground-truth endmember s_n and the estimate \hat{s}_n produced by each algorithm.

Table II shows median SAD over 50 realizations, using the original dataset and one with white Gaussian noise added to achieve $\text{SNR} = 25$ dB. In the noiseless case, the table shows that HUT-AMP recovered the grass and white TyVek materials with the highest accuracy and recovered the dry sand and black felt materials with the second highest accuracy. Meanwhile, in the noisy case, HUT-AMP recovered the TyVek material with the highest accuracy (in a tie with SDSNMF) and recovered the dry sand and black felt materials with the second highest accuracy. Looking at the material-averaged SAD scores in the table, it is evident that the accuracies achieved by HUT-AMP are close to those attained by the most accurate algorithm, SDSNMF, and significantly better than those attained by any of the competing algorithms, in both the noiseless and noisy cases. Although SDSNMF offers slightly more accurate endmember recoveries, Fig. 14 shows that its runtime is 44 times slower than that of HUT-AMP. Therefore we conclude that HUT-AMP offers an excellent combination of endmember recovery accuracy and runtime.

For visual comparison, Fig. 13 shows an example of the extracted and ground-truth endmembers in the noiseless case. The figure shows HUT-AMP’s estimates closely matching the ground-truth for all materials; by contrast, MVSA is mismatched in the case of grass, MVSA and FSNMF are mismatched in the case of dry sand, MVSA, VCA, and FSNMF, are mismatched in the case of black felt, and MVSA, VCA, SCU, FSNMF, and MVSA are all mismatched in the case of white TyVek. Figure 13 reveals that MVSA does not always yield non-negative endmembers estimates, which may account for its relatively poor performance in all but our first experiment from Sec. IV-A.

As another visual comparison, Fig. 14 shows an example of the recovered abundance maps in the noiseless case. We reason



Fig. 12. False-color image of the cropped scene of the SHARE 2012 dataset [57].

		grass	dry sand	black felt	white TyVek	avg.
noiseless	HUT-AMP	1.54	1.13	3.53	0.39	1.65
	VCA	1.58	2.20	11.01	2.09	4.22
	FSNMF	1.65	1.68	7.36	1.46	3.03
	MVSA	4.57	10.42	45.47	1.60	15.52
	SCU	2.69	2.48	32.10	1.19	9.61
	SDSNMF	1.86	0.71	2.85	0.40	1.45
SNR = 25 dB	HUT-AMP	1.60	1.03	3.68	0.44	1.69
	VCA	1.53	1.77	11.72	2.17	4.30
	FSNMF	1.58	14.47	4.39	1.65	5.52
	MVSA	4.52	11.16	48.05	1.58	16.33
	SCU	1.82	2.03	8.24	1.73	3.47
	SDSNMF	1.99	0.81	3.37	0.44	1.65

TABLE II
MEDIAN SPECTRAL ANGLE DISTANCE (IN DEGREES) BETWEEN
RECOVERED AND GROUND-TRUTH ENDMEMBERS IN THE SHARE 2012
EXPERIMENT.

that the best recoveries are the ones that are the most pure within the green, tan, black, and white regions of Fig. 12, given that great care was taken during data collection to keep each region occupied by a single material. Figure 14 shows that, in the case of dry sand and black felt, the abundances recovered by HUT-AMP were the most pure and, in the case of grass and Tyvek, the abundances recovered by HUT-AMP were among the most pure. The other Bayesian approach, SCU, yielded abundance estimates with much less purity, and we conjecture that was due to its priors being less well-matched to this highly sparse scene. Meanwhile, SUnSAL-TV (using both EE techniques) failed to recover the black felt material, which we attribute to its lack of a sum-to-one constraint.

Average runtimes are also reported next to each algorithm in Fig. 14. There we see that HUT-AMP’s runtime was 4-9 times slower than the EE-and-inversion techniques but 44 times faster than SDSNMF and 166 times faster than SCU, the other Bayesian technique.

We also ran HUT-AMP-MOS on the SHARE 2012 dataset and found that it correctly estimated the presence of $N = 4$ materials, thus yielding identical recovery performance to HUT-AMP. HUT-AMP-MOS’s runtime was 36.54 seconds, which was 2.5 times slower than HUT-AMP but still 67 times faster than SCU.

⁶The SHARE 2012 Avon dataset can be obtained from <http://www.rit.edu/cos/share2012/>.

⁷In practical HU data, ground truth is difficult to obtain, since lab-measured reflectivity can differ dramatically from received radiance at the sensor. In this experiment, we circumvent these problems by exploiting the known purity of the pixels and by minimizing noise effects through averaging.

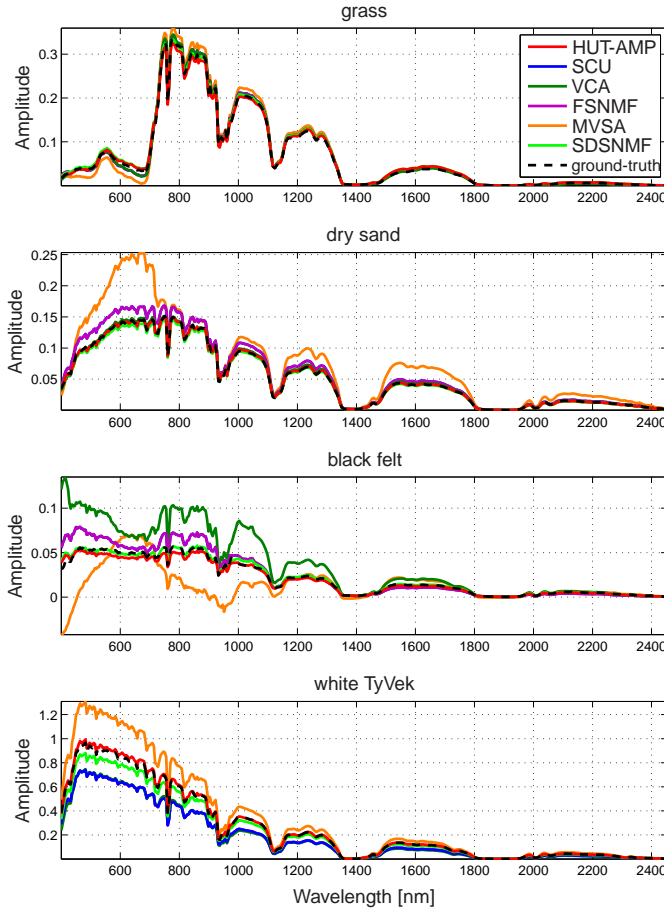


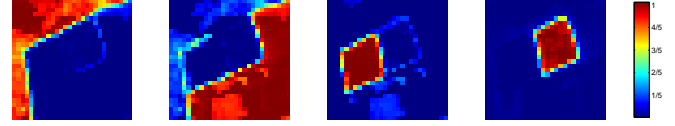
Fig. 13. Examples of recovered and ground-truth endmembers for the SHARE 2012 experiment.

D. AVIRIS Cuprite Dataset

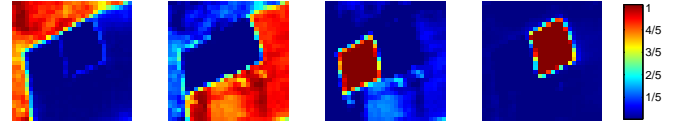
Our final numerical experiment was performed on the well known AVIRIS Cuprite dataset. Although the original dataset consisted of $M = 224$ spectral bands, ranging from 0.4 to 2.5 μm , we aimed to replicate the setup in [26], which removed bands 1-10, 108-113, 153-168, and 223-224 to avoid water-absorption effects, resulting in $M = 189$ spectral measurements per pixel. And, like [26], we considered only the 80×80 pixel scene identified by the black square in Fig. 15 and we assumed $N = 5$ materials. According to the tricorder classification map in Fig. 15, this scene contains the materials Montmorillonite, Alunite, well crystallized (wxl) Kaolinite, and partially crystallized (pxl) Kaolinite. Although [26] conjectured that this area also contains Sphene, none of the algorithms produced endmember estimates that were close to Sphene, and Sphene is not listed in Fig. 15. Thus, we did not consider Sphene as a ground-truth material. Also, like in [26], we considered both noiseless and white-Gaussian-noise corrupted measurements (at $\text{SNR} = 30$ dB).

Table III shows the median SAD achieved during endmember extraction over 50 realizations. From the table, we see that, in the noiseless case, HUT-AMP achieved the best material-averaged SAD as well as the best SAD for two specific materials. In the noisy case, HUT-AMP achieved the second-best material-averaged SAD as well as the best SAD for one

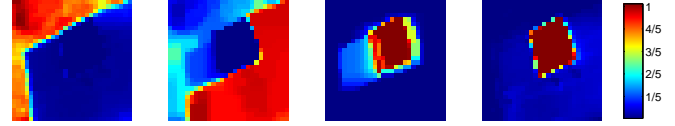
(a) HUT-AMP (average runtime = 14.68 sec):



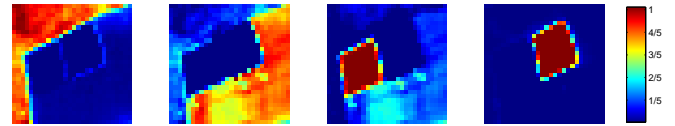
(b) VCA+FCLS (average runtime = 2.50 sec):



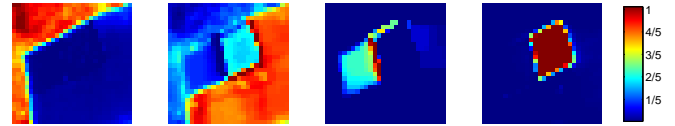
(c) VCA+SUnSAL-TV (average runtime = 4.13 sec):



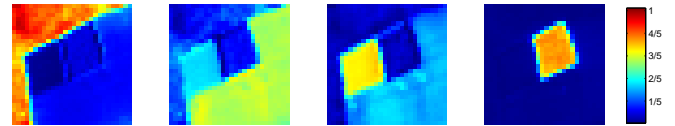
(d) FSNMF+FCLS (average runtime = 1.67 sec):



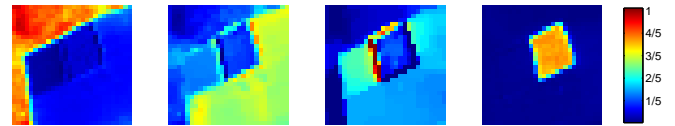
(e) FSNMF+SUnSAL-TV (average runtime = 3.36 sec):



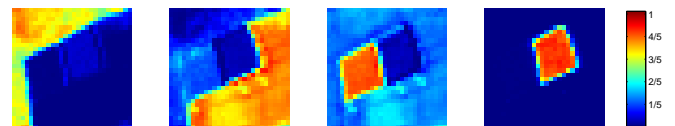
(f) MVSA+FCLS (average runtime = 1.82 sec):



(g) MVSA+SUnSAL-TV (average runtime = 4.08 sec):



(h) SCU (average runtime = 2438 sec):



(i) SDSNMF (average runtime = 642.8 sec):

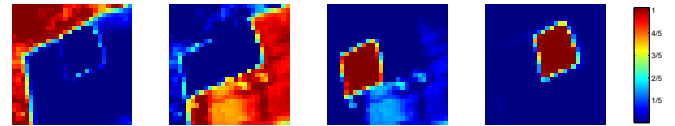


Fig. 14. Average runtimes and examples of recovered abundance maps for the SHARE 2012 experiment. From left to right, the materials are: grass, dry sand, black felt, and white TyVek.

material. Meanwhile, the SADs produced by VCA, FSNMF, and SDSNMF were of a similar magnitude, while those produced by SCU and MVSA were noticeably larger. These SAD values should be interpreted with caution, however, since i) the ground-truth endmembers are laboratory-measured reflectance spectra from the 2006 USGS library as ground-

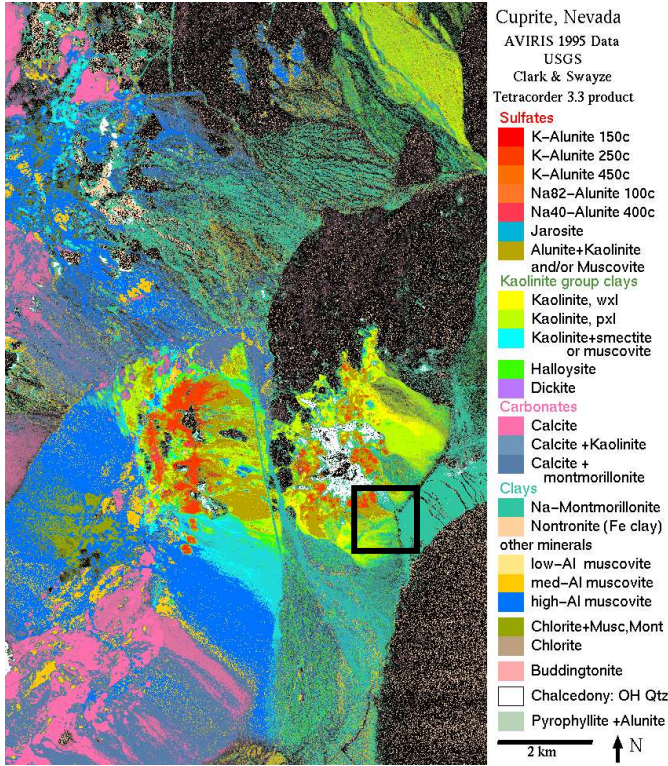


Fig. 15. Mineral classification mapping of the Cuprite Dataset using the Tricorder 3.3 product [59]. We used the scene cropped by the black rectangle.

		SAD [degrees]					
		HUT-AMP	SCU	VCA	FSNMF	MVSA	SDSNMF
noiseless	Montmor.	3.42	3.95	3.91	3.54	6.03	3.47
	wxl Kaolinite	10.23	13.14	10.45	10.86	15.42	11.46
	pxl Kaolinite	9.10	11.45	9.22	9.38	10.51	9.09
	Alunite	7.27	6.62	6.55	6.40	7.11	7.87
	Average	7.50	9.12	7.53	7.55	9.77	7.97
SNR = 30 dB	Montmor.	3.53	3.80	3.79	3.57	5.64	3.48
	wxl Kaolinite	10.72	12.46	10.62	12.93	15.59	11.20
	pxl Kaolinite	9.10	11.47	9.32	10.49	11.55	9.43
	Alunite	7.45	7.94	6.60	6.39	7.16	6.67
	Average	7.70	8.92	7.59	8.34	9.98	7.70

TABLE III
MEDIAN SPECTRAL ANGLE DISTANCE (IN DEGREES) FOR THE CUPRITE EXPERIMENT.

truth, whereas the Cuprite dataset itself uses reflectance units obtained via atmospheric correction of radiance data,⁸ and ii) it is not clear exactly which materials are truly present in the scene. The fact that the SADs reported here are so much larger than those reported in our SHARE experiment suggests that the Cuprite ground-truth may not be fully accurate.

For visual comparison, we plot examples of the abundance maps recovered in the noiseless experiment in Fig. 16. The figure shows that the abundance maps returned by HUT-AMP, SDSNMF, FSNMF+FCLS, VCA+FCLS, FSNMF+SUnSAL-TV, and VCA+SUnSAL-TV have the highest contrast, suggesting that if certain pixels are truly pure then these algorithms are accurately estimating those pixels. The maps

produced by SUnSAL-TV appear more “blurred,” probably as an artifact of TV regularization. The abundances returned by SCU, MVSA+FCLS, and MVSA+SUnSAL-TV were of much lower contrast and suggest different material placements than the maps generated by the other algorithms. For example, SCU suggests a significant wxl-Kaolin presence throughout the lower half of the scene, in contrast to other algorithms. However, Table III shows that SCU gave the worst SAD for wxl-Kaolin.

Figure 16 also shows the total runtimes of the various algorithms. There we see that HUT-AMP was 6-8 times slower than the typical EE-and-inversion approach, but more than 80 times faster than SCU and more than 200 times faster than SDSNMF.

We also ran HUT-AMP-MOS on the Cuprite data and found that, in both the noiseless and noisy cases, it estimated the presence of $N = 5$ materials, and thus returned identical estimates to HUT-AMP. Meanwhile, HUT-AMP-MOS gave an average runtime of 191.49 seconds, which was 30 times faster than SCU and 75 times faster than SDSNMF.

V. CONCLUSIONS

In this paper, we proposed a novel empirical-Bayesian hyperspectral-unmixing algorithm that jointly estimates endmembers and abundance maps while exploiting the practical features of spectral and spatial coherence, as well as abundance sparsity. Inference is performed using the “turbo” approach proposed in [34], which breaks up the factor graph into three subgraphs, performs (loopy) BP individually on each subgraph, and then exchanges beliefs between subgraphs. For the spectral and spatial coherence subgraphs, standard Gauss-Markov and discrete-Markov methods [28], [29], respectively, are used, while for the non-negative bilinear-mixing subgraph, the recently proposed BiG-AMP method from [30] is used, which exploits the approximate message passing framework from [31], [32]. Furthermore, the statistical parameters of all distributions are learned using expectation-maximization [40], and the number of materials in the scene is estimated using penalized log-likelihood maximization. On the whole, the proposed HUT-AMP-MOS algorithm performs approximate MMSE inference that exploits spectral and spatial coherence, in addition to simplex constraints, while avoiding the need for the specification of any tuning parameters.

Through a detailed numerical study, we demonstrated that our proposed HUT-AMP algorithm yields accurate recoveries of both endmembers and abundances on both synthetic and real-world datasets. In particular, we found that HUT-AMP gives recoveries that are close to—if not more accurate than—state-of-the-art unmixing algorithms like SDSNMF. Meanwhile, the runtime required for HUT-AMP is much less than sophisticated spatial-coherence exploiting approaches like SDSNMF and SCU—often by several orders of magnitude—while within an order of magnitude of the fastest EE-and-inversion approach. Our experiments also demonstrated that our model-order selection technique was able to correctly estimate the number of materials in several synthetic and real-world datasets, without requiring a very large increase in runtime.

⁸The reflectance and radiance versions of the Cuprite dataset can be found at <http://aviris.jpl.nasa.gov/html/aviris.freedata.html>

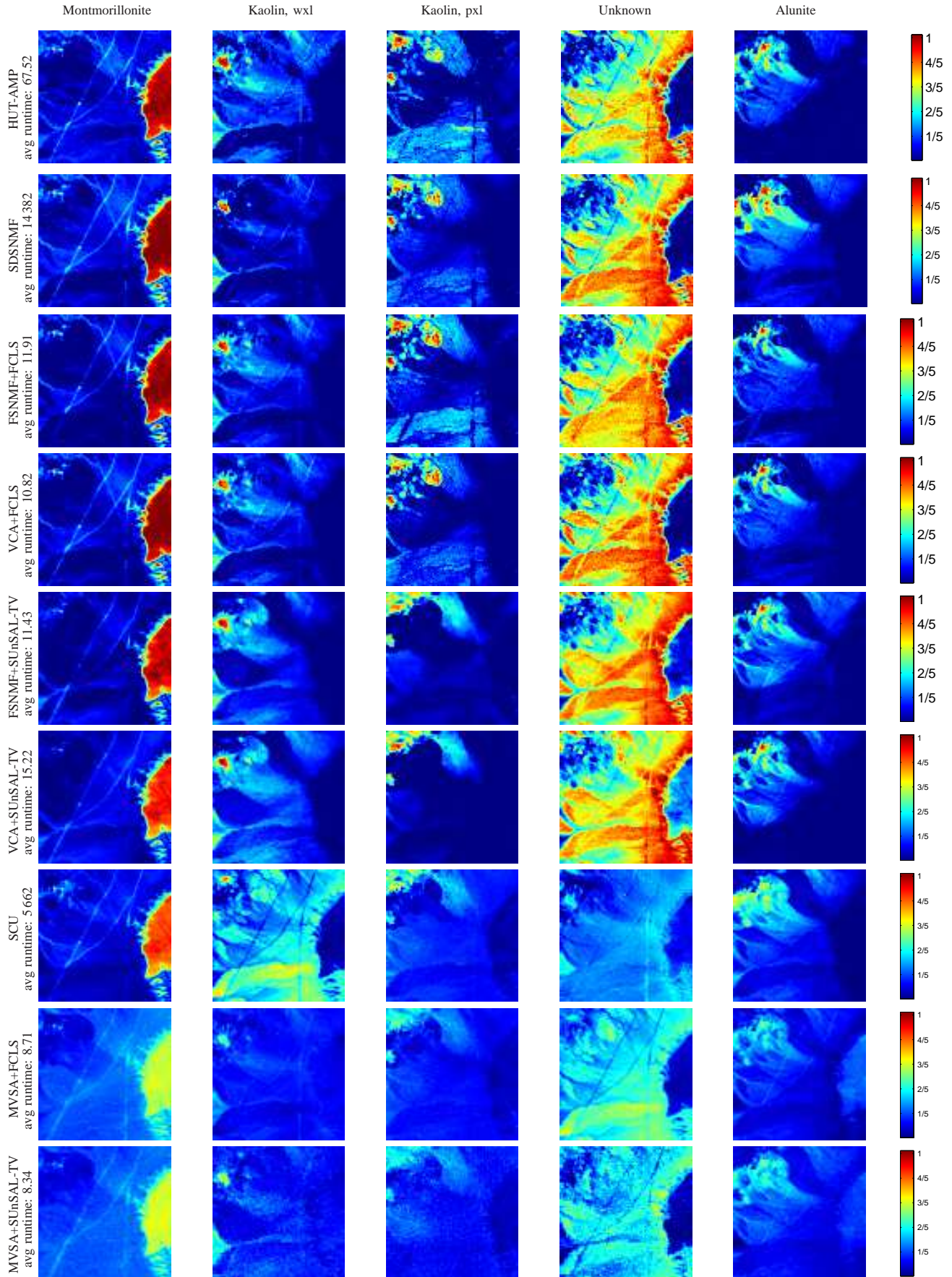


Fig. 16. Examples of recovered abundance maps in the noiseless Cuprite experiment. Each row corresponds to an algorithm and each column corresponds to a material. Average runtimes (in seconds) are also listed on the left.

APPENDIX A MEAN REMOVAL

We can see that \underline{S} from (7) is approximately zero-mean via

$$0 = \frac{1}{MT} \sum_{m=1}^M \sum_{t=1}^T y_{mt} \quad (50)$$

$$= \underbrace{\frac{1}{MT} \sum_{m=1}^M \sum_{t=1}^T \sum_{n=1}^N s_{mn} a_{nt}}_{O(1)} + \underbrace{\frac{1}{MT} \sum_{m=1}^M \sum_{t=1}^T w_{mt}}_{O(1/N)} \quad (51)$$

$$\approx \sum_{n=1}^N \underbrace{\frac{1}{T} \sum_{t=1}^T a_{nt}}_{\triangleq \mu_n^a} \frac{1}{M} \sum_{m=1}^M s_{mn}, \quad (52)$$

where (50) follows from the definitions (5)-(6). The underbraces in (51) show the scaling on each term in the large-system limit (i.e., as $N \rightarrow \infty$). These particular scalings follow from our assumption that the noise is both zero-mean and white and the convention [30] that both y_{mt} and the noise variance ψ scale as $O(1)$. Recalling that $\sum_{n=1}^N \mu_n^a = 1$ due to the simplex constraint, expression (52) shows that a weighted average of elements in \underline{S} is approximately zero, where the approximation becomes exact in the large-system limit.

REFERENCES

- [1] J. Vila, P. Schniter, and J. Meola, "Hyperspectral image unmixing via bilinear generalized approximate message passing," *Proc. SPIE*, vol. 8743, no. 87430Y, p. 9, 2013.
- [2] P. E. Johnson, M. O. Smith, S. Taylor-George, and J. B. Adams, "A semiempirical method for analysis of the reflectance spectra of binary mineral mixtures," *J. Geophysical Research*, vol. 88, pp. 3557–3561, 1983.
- [3] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 5, no. 2, pp. 354–379, 2012.
- [4] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 7, no. 6, pp. 1844–1868, Jun. 2014.
- [5] N. Dobigeon, J.-Y. Tournet, C. Richard, J. Bermudez, S. McLaughlin, and A. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 82–94, 2014.
- [6] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral endmember determination in hyperspectral data," *Proc. SPIE*, pp. 266–275, 1999.
- [7] J. Nascimento and J. Bioucas-Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, 2005.
- [8] D. L. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?" in *Proc. Neural Inform. Process. Syst. Conf.*, 2003.
- [9] S. Arora, R. Ge, R. Kannan, and A. Moitra, "Computing a nonnegative matrix factorization – provably," in *Proc. Sym. Thy. Computing*, 2012, pp. 145–162.
- [10] N. Gillis and S. A. Vavasis, "Fast and robust recursive algorithms for separable nonnegative matrix factorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 698–714, 2012.
- [11] J. Li and J. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 3, Jul. 2008, pp. 250–253.
- [12] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4418–4432, Nov. 2009.
- [13] D. Heinz and C.-I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, 2001.
- [14] R. Heylen, D. Burazerovic, and P. Scheunders, "Fully constrained least squares spectral unmixing by simplex projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4112–4122, 2011.
- [15] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.
- [16] —, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.
- [17] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imaging Vis.*, vol. 20, pp. 80–97, 2004.
- [18] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $l_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4282–4297, Nov. 2011.
- [19] Y. Yuan, M. Fu, and X. Lu, "Substance dependence constrained sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 2975–2986, Jun. 2015.
- [20] X. Lu, H. Wu, and Y. Yuan, "Double constrained NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2746–2758, May 2014.
- [21] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.
- [22] N. Gillis, D. Kuang, and H. Park, "Hierarchical clustering of hyperspectral images using rank-two nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2066–2078, Apr. 2015.
- [23] X. Liu, X. Xia, B. Wang, and L. Zhang, "An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 757–772, Feb. 2011.
- [24] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 161–173, Jan. 2009.
- [25] N. Dobigeon, S. Moussaoui, M. Coulon, J. Y. Tournet, and A. Hero, "Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4355–4368, 2009.
- [26] R. Mittelman, N. Dobigeon, and A. Hero, "Hyperspectral image unmixing using a multiresolution sticky HDP," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1656–1671, 2012.
- [27] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Proc. Uncertainty Artif. Intell.*, 1999, pp. 467–475.
- [28] C. A. Bouman, "Markov random fields and stochastic image models," in *IEEE Int. Conf. Image Processing Tutorial*, Oct. 1995.
- [29] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, 3rd ed. London: Springer, 2009.
- [30] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing—Part I: Derivation," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5839–5853, Nov. 2014.
- [31] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.
- [32] —, "Message passing algorithms for compressed sensing: I. Motivation and construction," in *Proc. Inform. Theory Workshop*, Cairo, Egypt, Jan. 2010, pp. 1–5.
- [33] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inform. Thy.*, Aug. 2011, pp. 2168–2172, (full version at [arXiv:1010.5141](https://arxiv.org/abs/1010.5141)).
- [34] P. Schniter, "Turbo reconstruction of structured sparse signals," in *Proc. Conf. Inform. Science & Syst.*, Princeton, NJ, Mar. 2010, pp. 1–6.
- [35] —, "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1462–1474, Dec. 2011.
- [36] M. Nassar, P. Schniter, and B. Evans, "A factor-graph approach to joint OFDM channel estimation and decoding in impulsive noise environments," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1576–1589, Mar. 2014.
- [37] S. Som and P. Schniter, "Compressive imaging using approximate message passing and a Markov-tree prior," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3439–3448, Jul. 2012.

- [38] —, “Approximate message passing for recovery of sparse signals with Markov-random-field support structure,” Jul. 2011, *Internat. Conf. Mach. Learning—Workshop on Structured Sparsity: Learning and Inference*, (Bellevue, WA).
- [39] J. Ziniel and P. Schniter, “Dynamic compressive sensing of time-varying signals via approximate message passing,” *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5270–5284, Nov. 2013.
- [40] J. P. Vila and P. Schniter, “An empirical-Bayes approach to recovering linearly constrained non-negative sparse signals,” *IEEE Trans. Signal Process.*, vol. 62, no. 18, pp. 4689–4703, Sep. 2014.
- [41] B. Efron, *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. New York: Cambridge University Press, 2010.
- [42] P. Stoica and Y. Selén, “Model-order selection: A review of information criterion rules,” *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.
- [43] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, “Factor graphs and the sum-product algorithm,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.
- [44] M. Bayati and A. Montanari, “The dynamics of message passing on dense graphs, with applications to compressed sensing,” *IEEE Trans. Inform. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.
- [45] S. Rangan, P. Schniter, E. Riegler, A. Fletcher, and V. Cevher, “Fixed points of generalized approximate message passing with arbitrary matrices,” in *Proc. IEEE Int. Symp. Inform. Thy.*, Jul. 2013, pp. 664–668, (full version at *arXiv:1301.6295*).
- [46] F. Krzakala, A. Manoel, E. W. Tramel, and L. Zdeborová, “Variational free energies for compressed sensing,” in *Proc. IEEE Int. Symp. Inform. Thy.*, Jul. 2014, pp. 1499–1503.
- [47] M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tourneret, A. Hero, and S. McLaughlin, “Tutorial on stochastic simulation and optimization methods in signal processing,” *arXiv:1505.00273*, 2015.
- [48] J. T. Parker, P. Schniter, and V. Cevher, “Bilinear generalized approximate message passing—Part II: Applications,” *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5854–5867, Nov. 2014.
- [49] Y. Altmann, M. Pereyra, and J. Bioucas-Dias, “Collaborative sparse regression using spatially correlated supports – application to hyperspectral unmixing,” *arXiv:1409.8129v1*, 2014.
- [50] G. F. Cooper, “The computational complexity of probabilistic inference using Bayesian belief networks,” *Artificial Intelligence*, vol. 42, pp. 393–405, 1990.
- [51] R. J. McEliece, D. J. C. MacKay, and J.-F. Cheng, “Turbo decoding as an instance of Pearl’s ‘belief propagation’ algorithm,” *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 140–152, Feb. 1998.
- [52] A. Dempster, N. M. Laird, and D. B. Rubin, “Maximum-likelihood from incomplete data via the EM algorithm,” *J. Roy. Statist. Soc.*, vol. 39, pp. 1–17, 1977.
- [53] J. P. Vila and P. Schniter, “Expectation-maximization Gaussian-mixture approximate message passing,” *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.
- [54] R. Neal and G. Hinton, “A view of the EM algorithm that justifies incremental, sparse, and other variants,” in *Learning in Graphical Models*, M. I. Jordan, Ed. MIT Press, 1998, pp. 355–368.
- [55] J. Bioucas-Dias and J. Nascimento, “Hyperspectral subspace identification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
- [56] C.-I. Chang and Q. Du, “Estimation of number of spectrally distinct signal sources in hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.
- [57] A. Giannandrea *et al.*, “The SHARE 2012 data collection campaign,” *Proc. SPIE*, vol. 8743, no. 87430F, p. 15, 2013.
- [58] K. Canham, D. Goldberg, J. Kerekes, N. Raqueno, and D. Messinger, “SHARE 2012: Large edge targets for hyperspectral imaging applications,” *Proc. SPIE*, vol. 8743, no. 87430G, p. 9, 2013.
- [59] R. N. Clark, G. A. Swayze, K. E. Livo, R. F. Kokaly, S. J. Sutley, J. B. Dalton, R. R. McDougal, and C. A. Gent, “Imaging spectroscopy: Earth and planetary remote sensing with the USGS tetracorder and expert systems,” *J. Geophys. Res.*, vol. 108, no. E12, p. 5131, 2003.